

The limits of sharing

President Clinton's offer to share missile defence with his allies brings to mind his presidency's weaknesses — his fondness of fudge and reluctance to embrace unpopular truths.

European leaders have been muted in their public responses to President Bill Clinton's offer, made in Lisbon last week, to share US missile defence technology with "other civilized nations". If this offer was meant to help bridge the growing chasm between European and American perceptions of the missile defence issue, it richly deserved to fail.

Reports that the United States has developed a working technology for national missile defence are grossly exaggerated. Sixty billion dollars have been spent since President Ronald Reagan first proposed such a system, but little real progress has been made. Space-based laser weapons and other fantasies have been jettisoned, and the Ballistic Missile Defense Organization (BMDO) at the Pentagon confines itself to the problem of intercepting a rocket by launching a mechanical interceptor in its path.

Despite this, Reagan's Star Wars fantasy has, since its early days, been accepted by a large part of the US population. People like to believe that American science and technology can solve any problem, and the Star Wars idea was a political success for Reagan. The fact that his vision has been technically discredited, and that tens of billions of dollars were spent with precious little to show for them, hasn't much diminished the political potency of missile defence.

So Reagan's heirs continue to pursue it. In its pursuit, they have even summoned up a category of enemy never previously encoun-

tered in the history of warfare — the 'rogue nation', whose leaders are not subject to the logic of intimidation. Opinion polls in the United States show a craving for national missile defence, and that many believe it works and even that it is actually already in place.

Last year, Clinton therefore found it politically necessary to declare that he would decide this summer whether to deploy a national missile defence system. The system would be the fruits of the BMDO, which spends 70% of its \$3 billion budget on battlefield defences, 20% on national missile defence and just 10% on technology development. Its technology development plan (see <http://www.acq.osd.mil/bmdo/bmdolink/html/tech.html>) confirms that the Star Wars dreams have been shelved. There is no real research any more, just a clumsy interceptor system that BMDO engineers are struggling to operate.

The first two formal tests of this system have been unsatisfactory: there are convincing allegations, not adequately refuted, that their results were rigged. Scientific experts in the United States have denounced the system as unworkable (see *Nature* **404**, 799; 2000). A third test, due next month, will not change that picture. The pressure to deploy now, ahead of November's election, is purely political.

Europe's tactfully restrained response to this unedifying shambles seems appropriate. Clinton's failure to convince Western Europe and Russia to collaborate gives the United States an opportunity to re-examine its rush to deploy a national missile defence system. ■

Gene therapy's trials

One major laboratory has closed because of a clinical trial's tragic outcome. But others need publicly to review their roles.

When Jesse Gelsinger, an 18-year-old Arizona man, died during a gene-therapy experiment last autumn, the scientific community first rallied to find out what went wrong. Then politicians and bureaucrats started looking for someone to blame.

Fault has come to rest on the University of Pennsylvania's Institute for Human Gene Therapy (IHGT), whose clinical-trials programme has now been terminated. The IHGT had failed to inform the US Food and Drug Administration (FDA) of adverse events experienced by other patients treated before Gelsinger, according to an FDA letter to the university. But other institutions, agencies and individuals may share some moral culpability: when oversight breaks down, everyone involved in it must examine their role.

The sense that this process is far from settled emerged at a recent Senate hearing. When officials from the Department of Health and Human Services, National Institutes of Health (NIH) and FDA were asked whether recommendations made in 1998 about clinical-trial oversight had been implemented, none responded definitively. Weeks earlier, the health department's Office of Inspector General pointed out that the 1998 recommendations had been largely ignored. Those recommendations spoke directly to many of the problem areas in the Gelsinger trial: informed consent, adverse-events reporting and clinical-trial oversight. The agencies' failure to address issues raised years earlier may have indirectly contributed to the trial's tragic outcome.

Adverse-events reporting may be the clearest example of this.

Hundreds of previously unreported adverse events poured in to the NIH once it asked for them following Gelsinger's death. Many gene-therapy researchers privately object to filing adverse-events reports either to the FDA, where they remain confidential, or to the NIH, where they are made public. But, in retrospect, having more public records on adverse events associated with the vector used in Gelsinger's trial would have been useful. Perhaps if enough data on immune responses associated with the vector had been made public before his death, rather than after, the trial would never have been launched.

So whose fault is it that many gene-therapy investigators didn't report their adverse events earlier? The answer is hard to determine, but the NIH leadership may have sent an inadvertent message that reporting to the Recombinant DNA Advisory Committee (RAC) isn't compulsory when in 1996 they tried to reduce the committee's scope.

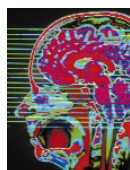
And although IHGT investigators didn't notify the FDA of all adverse events immediately, they did notify them of similar immune responses. In each case, the FDA let the trial resume. The IHGT and FDA changed the route of vector administration, from intravenously to directly into patients' livers, without notifying the RAC. While that adjustment was intended to limit the vector to the liver, Gelsinger's autopsy showed that it had the opposite effect.

Clearly, the IHGT's closure does not obviate the need for other institutions, agencies and individuals to assess their roles in the tragedy and to make their conclusions public. ■





Hard cheese
Final sequencing of
the mouse genome
'could be delayed'
p602



Getting ahead
Gateway to brain
data promised for
the Internet
p603



Hard choice
Spain faces dilemma
over whether to join
French synchrotron
p604



Ruled out
Tough new
regulations for
xenotransplants
p606

Biologists challenge sequencers on parasite genome publication

Colin Macilwain, Washington

Rival approaches to the way in which genome sequences are published are creating growing tension between scientists at the sequencing centres and those who want to use the sequencing data to further their study of the organisms involved.

International consortia that are sequencing *Plasmodium falciparum*, a parasite that causes malaria, and *Trypanosoma brucei* (see right), which causes sleeping sickness, are each currently engaged in heated arguments over the wisdom of publishing preliminary, annotated sequences in advance of the completion of full sequences.

In each case, prominent biologists who specialize in the organism are pushing for early publication. But genome sequencing centres say the publication of preliminary data could deprive their teams of the proper credit for the full, annotated sequence when it is completed.

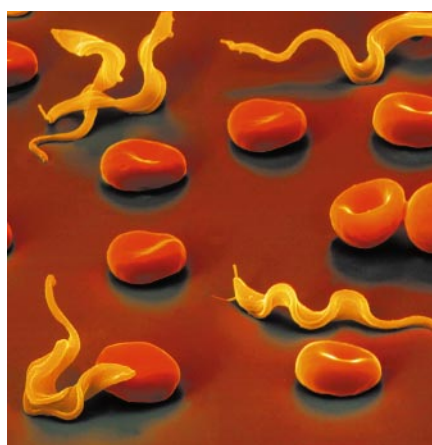
Claire Fraser, president of The Institute for Genomic Research (TIGR) at Rockville, Maryland, which is part of both consortia, believes that if outside scientists publish preliminary annotations (proposed function of a stretch of DNA) based on raw sequencing data made available voluntarily by the sequencing centres, the final complete sequence may never be published.

"It would be unlikely that the sequencing groups would come up with enough data, over and above those already published, to warrant publication," she says. "So it becomes not just an issue of credit for publication, but an issue of data piracy."

The problem Fraser describes is not confined to teams working on parasites involved in tropical diseases. Dozens of projects are under way to sequence the full genomes of interesting organisms, and most expect to take between three and five years to sequence, annotate and publish the complete genome.

The issue is common to most genome projects, says Michael Gottlieb, programme officer for parasitology at the National Institute of Allergies and Infectious Diseases, which is supporting work in both consortia.

"Our concern is to achieve an appropriate



Deadly sequence: *Trypanosoma brucei*, the protozoan responsible for sleeping sickness.

balance between the community's interest in getting the information out as soon as possible, and the sequencer's interest in getting the first opportunity to publish the annotated sequence," he says.

The small communities of researchers specializing in these organisms want information about the organisms' genes as quickly as possible. And although the sequencing centres usually release the raw sequence data

on the Internet as they obtain them, these data are of limited use to microbiologists who lack the bioinformatics capacity needed to interpret them. Some are teaming up with bioinformaticists to partially annotate the public data themselves. The trouble starts when they try to disseminate these partial annotations to their colleagues.

The issue came up last November at the regular meeting of the *T. brucei* consortium, which rejected a request from George Cross of Rockefeller University in New York to publish an annotation of part of the genome. "We're not allowed to make the knowledge we have developed available to anyone else in the field," complains Cross.

"There is a tremendous and increasing tension" between the sequencing centres and the microbiologists, he claims. "The key issue is that the sequencing centres want the sequence to be 99.9 per cent complete, but most biologists can get tremendous input from interim data."

The same issue was set to arise in Britain this week at a meeting of the malaria genome consortium at the Sanger Centre in Cambridge. Lou Miller, a microbiologist at the National Institutes of Health in Bethesda, Maryland, was expected to ask to publish a

New Russian science head named

Carl Levitin, Moscow

Ilya Klebanov, a deputy prime minister with close links to Russia's military industries and its arms-exporting companies, has been appointed the country's new head of industrial and scientific policy. As such, Klebanov will control the new Ministry for Industry, Science and Technologies, which 'swallowed' the former Ministry of Science and Technologies.

Former science minister Mikhail Kirpichnikov says the decision to found the super-ministry (see *Nature* 405, 384; 2000) was taken "virtually at the moment of signing the presidential decree on the new

cabinet", and neither he, Yuri Osipov — the president of the Russian Academy of Sciences — nor Klebanov were consulted.

Kirpichnikov has expressed his concern to prime minister Mikhail Kasyanov about whether the new cabinet structure is the best way of managing Russian science. He says he was told this was a chance to unite scientific research 'scattered' over many programmes.

Osipov says Aleksandr Dondukov, the new minister of industry, science and technology, has promised to consult the academy before taking decisions on basic science, and hopes to build a powerful 'scientific unit' within the ministry.

▶ preliminary annotation of the *P. falciparum* sequence he has prepared with Eugene Koonin of the National Centre for Biotechnology Information.

Miller's proposal was met with indignation at the sequencing centres when he approached them in March. "What he's done has clearly gone against the interests of the funding agencies and sequence centres," says Fraser. "The sequencing groups felt they were being held hostage," she says, because Miller was offering to publish in collaboration with them, "but saying he would do it anyway".

Fraser says Miller backed off once the sequencing groups' hostility became apparent. Miller declines to speak on the record about his plans. He says public discussion is unlikely to improve relations between the parties involved. However, it is understood that he has no plans to publish his annotation without the consent of the consortium.

"He's not doing it to get the credit," says Malcolm Gardner, who heads the *Plasmodium* sequencing effort at TIGR. "He simply believes that it is in the best interests of the community that the information gets out there." However, Gardner adds, "people who have invested four years of work in this should have the privilege of publishing it."



Malaria: lethal in children.

the consortium two years ago.

"I ran into anxiety, sometimes bordering on paranoia, at the sequencing centres," he recalls. Roos published some important genes from the parasite in the *Proceedings of the National Academy of Sciences*. "At the time it caused a tremendous uproar. There was anxiety that we'd skimmed the cream from the project — but in fact what we did added value to what the sequencing centres do."

Roos thinks Miller's work should be included in the Internet portal his team is creating to make the malaria parasite sequence accessible to microbiologists. A first version of the portal will go live this week, at <http://e2kroos.cis.upenn.edu/PlasmoDB.html>.

David Roos, a microbiologist at the University of Pennsylvania, Philadelphia, has obtained funds from the Burroughs Wellcome Fund to present a preliminary annotation of the malaria parasite sequence on the Internet. But he says he encountered some hostility when he first became involved with

e2kroos.cis.upenn.edu/PlasmoDB.html.

Some researchers, however, will continue to press for publication of preliminary annotations of organism genomes. Most — but perhaps not all — will adhere to the data-release policies, posted beside the preliminary sequence data, which tell researchers to obtain permission before publishing results based on the data.

Separate genome projects will consult leading scientific journals about whether the release of preliminary annotation on the web will prejudice subsequent publication of complete, annotated chromosomes.

But ultimately these tensions could change the way biology is published. Roos points out that high-energy physicists, who publish in teams of hundreds, are named on papers in alphabetical order, with no lead author. "My prediction is that the same will happen in genomics," he says.

Roos sits on a couple of departmental appointment boards at his university, and jokes that they'll have to start assessing candidates by their true contribution, rather than inferring it from the order in which their name has appeared on papers. "We'd need to think about what they've done," he says. "And that wouldn't be such a bad thing." ■

Drive for more genomes threatens mouse sequence

Alison Abbott

Mouse geneticists are expressing concern that completion of the mouse genome sequence could be delayed because of increasing pressure to sequence the genomes of other animals. The doubts were voiced at a recent workshop organized by the National Institutes of Health (NIH) to set priorities for resources for mouse genetics and genomics.

The mouse was selected two years ago as the second mammalian species, after humans, to have its genome sequenced both by a public consortium led by the NIH and by the private company Celera. This is because the mouse is widely considered to be the best-understood species genetically.

The number of mutant mouse strains is growing rapidly. These are being generated using technologies that 'knock out' targeted genes and by random chemical-mutagenesis screens, where mice are exposed to a mutagen and those displaying interesting new traits are selected for breeding. Many of these mutant strains are potentially important models for human disease, and accurate information about their gene sequences will help to unravel the molecular bases of these diseases.

Within the past year, the NIH has set up 10 centres to sequence the mouse genome, and plans to divert capacity in human genome sequencing centres over to the mouse as the human genome nears

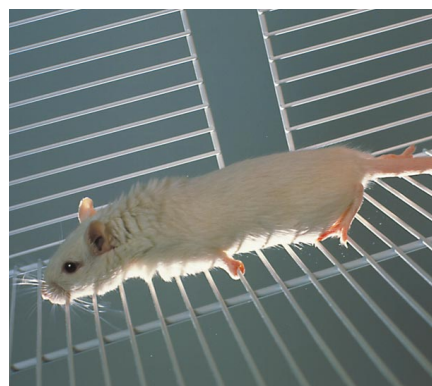
completion. But many other species, from the zebrafish to the rat, are also vying for the attention of sequencers.

"We always understood that the mouse genome would be finished, but learnt only at the workshop that a firm decision had been taken only for a working draft," says Rudi Balling, director of the Institute for Mammalian Genetics at the National Research Centre for Environment and Health in Munich.

So far, the mouse sequence is more advanced in the private than in the public domain. Last week Celera announced that after only two months it has sequenced more than a billion base pairs of the 129/SvJ mouse strain, the strain most often used to generate knockouts. This is estimated to be around one-third of the full genome.

Celera's president, Craig Venter, says the company will have the finished, assembled sequence by next summer. Correct assembly of the vast number of base pairs will be relatively easy, he says, as it can be directed by the blueprint of the human genome. This 'humanized mouse genome', as Venter calls it, will help to pinpoint genes on the human genome and help to identify gene-regulatory areas, he says.

But Celera's mouse genome sequence will be restricted to the company's subscribers. The public consortium, with its slower but



Second choice: the complete mouse genome sequence may be forced to take a back seat.

more exact approach, is aiming for a 'working draft' of the genome of the C57BL/6 mouse strain — widely used in genetic and immunological studies — well before its original target date of 2003, says Elke Jordan, deputy director of the NIH's National Human Genome Research Institute.

She declines to put a date on finishing, saying that in principle the NIH's commitment to finish "remains unchanged: it's what everybody wants". But she says there are "too many uncertainties about funding" for a date to be fixed.

This view is echoed by Eric Lander, director of the Whitehead Institute at the

► Massachusetts Institute of Technology. "There is tremendous enthusiasm from all quarters to finish the mouse genome, and scientists won't have to wait very long," he says.

But he admits that "the importance of a lot of different projects will have to be juggled, and it'll all get done in an order that is maximally useful to scientists as a whole". Like the NIH, Germany and France are waiting to assess the scientific promise of different species before committing their national sequencing efforts to finishing the mouse.

Jordan points out that the public consortium, concurrent with providing a working draft, will finish selected areas of the mouse genome judged to be biologically important. Although primarily intended to support priority areas of mouse genetics, this "will also give us a better idea of how much better it really [would be] to have the finished sequence", she says.

Mouse geneticists need no such convincing. "A draft sequence will be a poor tool for mouse geneticists," says Phil Avner, head of the mouse molecular genetics unit at the Pasteur Institute in Paris. It will allow no more than 80 per cent of genes to be identified, and will not provide key information about areas of the genome that regulate the genes, he says.

Even some zebrafish geneticists agree that finishing the mouse genome is a high priority because of the uniquely advanced level of genetic tools available. "But zebrafish biologists will also need a working draft within the next year or two," says Wolfgang Driever, a zebrafish geneticist at the University of Freiburg. "So ultimately it will be a question of timing and funding." ■

Internet gateway planned for neuroinformatics data

Paul Smaglik, Washington

Data on the human brain will soon be available over the Internet via an electronic gateway. Currently being planned by an international consortium, the portal will give researchers access to data at various levels of detail and sophistication.

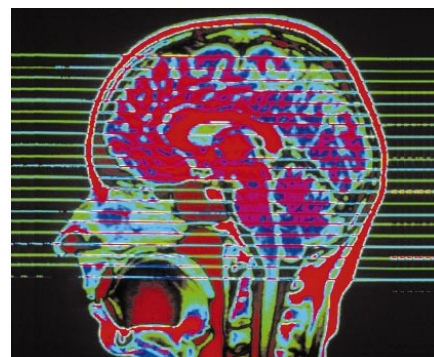
Backed in principle by the Organisation for Economic Cooperation and Development (OECD), the consortium aims to devise common standards and procedures for the various neuroinformatics databases and software programs scattered around the world.

An OECD working group set up to develop the gateway met for the first time in Genoa last month, and announced that it plans to make decisions on a launch date at its next meeting in Washington in September.

"We'd like to create a portal that would provide access to all the resources that are out there," says Stephen Koslow, director of the Office on Neuroinformatics with the US National Institute of Mental Health, and chair of the OECD working group.

Koslow points out that many individual programs already exist, each working at their own levels of analysis. Finding them and linking them to one site will be the first step, he explains. The next step will be to make them work with each other.

The project resembles earlier efforts to build computational tools for studying genomes. Many groups independently creat-



Head first: brain data, such as this MRI scan, will soon be brought together on the Internet.

ed computer programs that eventually became consolidated into software suites.

But neuroinformatics covers a broader spectrum of tools than computational genomics, ranging from brain imaging to the study of genes and proteins. And tying together data dealing with different levels of analysis is a huge challenge. Data can be in many forms, including brain scans showing the development of Alzheimer's disease and genetic databases detailing susceptibility to the disease.

Scientists involved in the Human Brain Project — a US government-backed effort to develop a variety of neuroinformatics tools — are already working to make the various pieces of software more accessible.

Jonathan Cohen, professor of psychology at Princeton University, said last week that although independence in the early days of neuroinformatics software development had resulted in many useful programs, it had also created an electronic 'Tower of Babel'.

"Since a lot of people build these things on their own, they are not always in a format that everyone can use," Cohen told a conference on the Human Brain Project held last week in Bethesda, Maryland.

Cohen's group is providing 'wrappers' to make different programs look the same on a computer screen. So far, this 'FisWidgets' project has provided interfaces for 43 public-domain neuroimaging programs.

Another problem is that data can be in different forms, or images in different resolutions. This is especially daunting when developing multiple-scale models of the nervous system, Nigel Goddard, a bioinformatician at the University of Edinburgh, told the meeting. Such modelling provides a broad map of a system, along with the ability to zoom in on selected parts for greater detail, and retrieve data associated with that image. "We need to have some set of standards so we can integrate the efforts," said Goddard. ■

Software spend boosts Israeli R&D

Haim Watzman, Jerusalem

A recalculation of Israel's spending on research and development (R&D) shows that it spends 3.5 per cent of its gross domestic product (GDP) on civilian research — more than any other OECD country. The figures are the first in Israel to include accurate data on spending by small software companies in Israel.

The new picture comes from a survey carried out by Israel's Central Bureau of Statistics. Previous assessments had placed this figure at just under 3 per cent of GDP (military R&D figures remain classified). Such estimates for high-tech R&D spending were based primarily on information from large companies, explains Simcha Bar-Eliezer, the statistician responsible for the survey.

Shlomo Herskovic, the director of planning and information for the Planning and Budgeting Committee of the Council

for Higher Education, which oversees and allocates the country's higher-education budget, says the new figures change the overall picture of Israel's spending on research.

The survey shows that 21 per cent of the country's civilian research spending is at the universities. Herskovic says the data could be used to argue that the universities should get more research funds. Given the extra spending in software companies, the government is now responsible for only 8 per cent of civilian research expenditure.

The high level of research spending in the software sector could also have implications for computer science programmes at universities. Even current plans to increase the number of computer science graduates by a factor of five over the next three years may be inadequate to supply software firms with the professionals they need, says Herskovic. ■

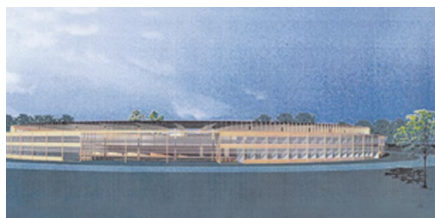
Spain in quandary over French synchrotron

Xavier Bosch, Barcelona

It has only been in existence for a few weeks, but the Spanish Ministry of Science and Technology is already facing its first major dilemma. The French government has invited Spain to participate in the funding and operation of France's proposed synchrotron, Soleil — but such a move could dash Spanish scientists' long-held hopes for their own machine.

At a meeting two weeks ago in Santander, Spanish and French science officials set up a working group on scientific issues. In particular, the group will address space policy and large scientific equipment, such as the 10-m Gran Telescopio de Canarias (Grantecan) project and the Soleil synchrotron.

Ramón Marimón, state secretary for science and technology policy, says that Spain will take the invitation to participate in Soleil seriously, and is keen to know where the facility would be built. Some observers argue that Spain is more likely to participate in the project if Soleil is built in a southern region of France near to Spain, rather than near Paris.



Hot topic: Spain must decide whether to join in with the planned French synchrotron, Soleil.

Marimón says French officials at the Santander meeting were told that the Ptas14 billion (US\$79 million) Grantecan has so far been financed by "national funds only". It is thought that Spain might seek French funding for the telescope — whose foundation stone was laid on 2 June — in return for any contribution to Soleil.

But Marimón says that the Spanish synchrotron, which the government has so far refused to finance, has not been ruled out. In December 1997, a design study was put forward for a 2.5–3 GeV synchrotron light facility called LLS (Laboratori Llum Sincrotró), to be housed at the Autonomous University of Barcelona in Catalonia.

The Catalan government offered to cover one quarter of the estimated Ptas13 billion cost, if the national government paid a further quarter, and the European Commission the rest. But the former Office of Science and Technology turned this down in January 1998.

Ten months later, the European Science Foundation (ESF) released a report on the needs for a European synchrotron and related beamlines for biological and biomedical research. Among the medium-term needs, the review panel recommended the construction of new sources, including the LLS.

Enric Banda, ESF general secretary, says the LLS is a "very good project" and adds that, to sort out the north–south imbalance, Europe should have a synchrotron in the south. "If there were to be a facility in southern Europe, Barcelona would be a logical place," he adds.

Andreu Mas-Colell, head of the Catalan government's Department of Research,



POKEMON CRAZE HITS EUROPE..

says that although the National Plan on Research, Development and Technological Innovation does not rate a synchrotron as a priority, his department sees a need for a big physics installation. "Hence the LLS is a priority," he says.

Marimón says the Catalan government's funding formula "seems reasonable". Much like Mas-Colell, LLS director Joan Bordas believes that the traditional scientific rivalry between Madrid and Barcelona may have blocked the project so far. "I don't mind if the facility is built in another Spanish region," he says. "But Spain does need a third-generation synchrotron radiation source."

Grantecan, which the national research plan has prioritized (see *Nature* 400, 393; 1999), must get 30 per cent of its funding from abroad, but no foreign partner has yet been found.

Immigrants help offset Canada's brain-drain crisis

David Spurgeon, Montreal

For years Canada has complained of a 'brain drain' of its best and brightest, lured by higher salaries and better career prospects in the United States. But, according to the country's Liberal government, the latest statistics indicate that it might be experiencing more 'gain' than 'drain'.

An article in the May issue of *Education Quarterly Review*, a publication of the Canadian federal body Statistics Canada, says the country gained four university graduates from abroad for every one it lost to the United States. Furthermore, Canada received as many immigrants with a master's degree or doctorate as it lost university graduates at all levels to the United States.

The analysis confirms the widely held perception that the loss of highly skilled workers to the United States accelerated during the 1990s. But so too did the influx of

such individuals from abroad: about one-third of the increase in employment among computer engineers, programmers and systems analysts during the decade consisted of immigrants, many coming from developing countries.

During the 1990s, between 22,000 and 35,000 people a year are estimated to have emigrated from Canada to the United States, either temporarily or permanently. They were more likely to be graduates and have higher incomes than the Canadian population as a whole. Analysis of income tax data showed that individuals earning more than Can\$150,000 (US\$100,000) a year were seven times as likely to leave as the average taxpayer.

Most of those who left were in the 23–44 age group. Many worked in hospitals, education and high-technology industries such as telecommunications and services, precisely the 'knowledge-intensive' sectors

that the federal government has emphasized as important to the Canadian economy.

But from the mid-1980s to 1997, the most recent year for which statistics are available, permanent immigration to Canada increased 15-fold among computer scientists, 10-fold among engineers, eightfold among natural scientists and fourfold among managerial workers. In 1997, more than 20,000 immigrants intended to work as computer scientists, engineers and natural scientists.

Canadian business leaders have often complained about the lack of recruits for the country's hi-tech industries. Thomas d'Aquino, chief executive officer of the Business Council on National Issues, says the study does not show that Canada has no problem. "If we didn't have a brain gain from immigration, we'd be in really serious trouble," he says.

A slot in the dock predicted for the chemical industry

Natasha Loder, London

The chemical industry could be facing a future filled with liability claims sparked by increased knowledge of the human genome, according to a British environmental group.

A report by Friends of the Earth, *Crisis in Chemicals*, warns that knowledge derived from the Human Genome Project on how chemicals affect individuals could open the legal floodgates. The group also calls for new legislation on the toxicity of chemicals.

New genomics data will mean that the chemical industry will be able to look at the human effects of a new product before it is made available. But it also means that companies failing to do so might face legal liabilities for negligence.

Industry, governments and the financial sector need to recognize the full implications of the biomedical revolution, says Michael Warhurst, one of the report's authors. "Companies that fail to clean up their act will face a heavy bill," he says. "So will insurers and investors."

The UK government has welcomed the report. "It highlights the possible implications of emerging research into the genetic susceptibility of individuals to chemicals, and rightly points out that this research will eventually bring a better understanding of the mechanisms of chemical toxicity," says a spokesperson.

Friends of the Earth predicts that, without action now, there will be a crisis in the regulation and use of chemicals. The UK government is due to hold a 'stakeholder forum' on the regulation of chemical use in the next few months, where this will be discussed and which Friends of the Earth says it will attend.

But Warhurst says that Europe is where the real action on chemical regulation is taking place, and the European Commission is currently reviewing the regulation of industrial chemicals. A UK government spokesperson says: "The report makes some interesting recommendations for the future of European chemical legislation which we will be studying carefully."

The Chemical Industries Association, a lobbying umbrella group for the UK chemical industry, has responded to the report by saying that a programme of "product stewardship" currently ensures that "nobody is exposed to danger from our products". It adds that the chemical industry "is extremely highly regulated". ■

► <http://www.foe.co.uk>

► <http://www.cia.org.uk>



End of an era: funded as an experiment for 30 years by Roche, but left free to direct its own research, the Basel Institute for Immunology is set to close down.

Roche brings down curtain on Swiss immunology lab

Alison Abbott, Munich

The Swiss-based pharmaceutical company Roche announced this week that it is to close its prestigious Basel Institute for Immunology (BII), which has been home to three Nobel prizewinners. The institute will be transformed into a medical genomics centre under the interim directorship of Klaus Lindpaintner, head of Roche's genetics activities in Europe.

The move brings to an end the 30-year experiment of a research institute being supported in its entirety by a pharmaceutical company, and given complete academic freedom to pursue any line of research in immunology.

The BII's 160 employees are in shock. Although Roche had indicated three years ago its waning interest in supporting the institute, discussions had never been sufficiently intense to suggest that such a radical move was on the near horizon, says Fritz Melchers, BII's director for the past 20 years.

Roche has enjoyed significant reflected glory from the BII, which cost SFr40 million (US\$24 million) per year to run. But it has not enjoyed any profit, and has not picked up a single lead from the institute's research. "The model Roche invented was never taken up by other companies," points out Jonathan Knowles, head of Roche's global research organization.

Knowles says that "it was appropriate to support the BII 30 years ago, when the worlds of academia and the pharmaceutical industry were strictly separate". But times have changed. "Academics and industry are now happy to work together synergistically," he says. The company will work much more closely with the new Roche Center for Medical Genomics than it did with the BII, where he says that a lack of dialogue had become an issue of concern to the company.

Knowles says that the model of operation will be similar to that of the Glaxo Institute for

Molecular Biology in Geneva, at which he was director until joining Roche two years ago.

The Center for Medical Genomics will focus its activities on common diseases with complex genetic bases, such as cardiovascular disease, cancer, and psychiatric and neurodegenerative disorders, using human material rather than model organisms.

Roche has pledged to help all the staff find new jobs either within or outside the new institute. Nearly all staff are on five-year contracts, and these will be honoured.

Melchers is distressed by Roche's decision with which he, along with the international scientific advisory board, disagrees. Having done what he feels he could to help ensure the continuation of the institute, he appears to accept defeat philosophically. "Ultimately a company must be able to make its own decisions about its directions," he says. "I may not agree with this decision, but I respect it."

He is also disappointed about the failure of Roche's attempts, encouraged by the advisory board, to transfer the institute to a local university. The universities could not afford the running costs.

Nobel laureate Robert Huber, a director at the Max Planck Institute for Biochemistry in Munich, and a member of the BII's scientific advisory board, says that "the BII has been a world-leading institute and its closure is very, very sad". He adds: "But in retrospect I wish there had perhaps been a closer dialogue between the institute and the company — then it might not have come to this."

Manfred Eigen, of the Max Planck Institute for Biophysical Chemistry in Göttingen, is another of the advisory board's Nobel laureates, and a member of its board of directors. He says that the board chose to dissolve itself when it learnt of Roche's decision last week, rather than endorse a decision that it believed to be wrong. "But of course Roche fully financed the BII, and thus has the right to do what it wants." ■

US decides close tabs must be kept on xenotransplants...

Declan Butler

It may not be much of a silver lining. But recent US experience with clinical trials of gene therapy appears to have helped persuade the government that trials of transplants from animals to humans — xenotransplants — should be subject to strict federal regulation.

The move, outlined under government guidelines published on 25 May, acknowledges the risk that xenotransplants might introduce virulent viruses into humans. It represents a significant tightening of previous guidelines issued by the Public Health Service (PHS) four years ago.

The PHS produced the new draft after discussion among its component agencies: the Centers for Disease Control and Prevention, the Food and Drug Administration (FDA), the Health Resources and Services Administration, and the National Institutes of Health (NIH). The draft states that “all xenotransplant products pose a risk of infection and disease to humans” and that “all species pose infectious disease risks”.

Earlier guidelines would have left approval of clinical trials to local university review boards. The new draft recommends

instead that all clinical trials are approved by the FDA, and that xenotransplantation be overseen by a new committee, the Secretary's Advisory Committee on Xenotransplantation (SACX; see opposite).

“It is another strong step in making sure that any clinical trials are done in a fashion that is consistent with national policies rather than local views, which may be subject to influence or lack of critical expertise,” says Dan Salomon, a researcher at the Scripps Research Institute, former head of the American Society of Transplant Surgeons, and a leading xenotransplant proponent.

The recent controversy over the reporting of side effects in gene-therapy trials has helped the FDA withstand the current anti-regulatory climate in the United States, says Jonathan Allan, a virologist at the Southwest Foundation for Biomedical Research in San Antonio, Texas, and a member of the FDA's advisory subcommittee on xenotransplantation. The FDA “now feels it can regulate the field of xenotransplantation”, he says.

The earlier draft would have allowed transplants from non-human primates, despite the virological risk. The new document reaffirms an earlier FDA moratorium on such trials.



No monkey business: a ban on transplants from non-human primates has been retained.

Allan, who has long campaigned against transplants from non-human primates, agrees that the guidelines “are a significant improvement”. But he is concerned that the PHS position remains ambiguous, suggesting that the SACX “make recommendations on the questions of whether and under what conditions the use of non-human primate xenografts would be appropriate”.

There is provision for imported animals to be used when the “species or strain are not available for use in the United States and their use is scientifically

► warranted". Designed to allow companies such as Novartis to import transgenic pigs, this loophole could lead to use of virus-laden wild animals, says Allan.

The document does not go into detail on the viruses that need to be considered in risk assessment, or species-specific differences, leaving this for the FDA to consider for each protocol. Difficulty in assessing the risk is the major caveat to proceeding. Although the guidelines emphasize the need to assay tissues for viruses, in reality most new viruses are only detected after the event.

Most research on pig viruses has focused on those that cause losses to pig farmers. Little is known about viruses, such as herpes and retroviruses, that cause low-lying infections that might be dangerous to humans.

"I still worry about the infectious disease risks," says Allan. "But if there are promising therapies, such as pig neuronal cells to treat Parkinson's disease, that could benefit millions it changes your notion of how to go forward." Such cells may be less dangerous than whole-organ transplants. But more work is needed to develop primate models for viral infection.

Abdullah Daar, a surgeon and bioethicist at Sultan Qaboos University in Oman, says the new guidelines are "much tougher", but is sceptical as to how they will be implemented. They require sponsors of trials to ensure informed consent of patients, their families and close contacts, for example, and long-

term surveillance of subjects. But several observers doubt that this is feasible.

The PHS also recommends that blood and tissues samples and all records of trials be kept for 50 years. Salomon wonders who will pay for such a proposed national database: "Industry will be reluctant, and there is certainly no evidence that the FDA could budget such a project."

"The PHS is determined to go ahead with clinical trials, but if we are putting the

public at risk then broad public consultation is needed, and not decisions by experts," says Fritz Bach, a xenotransplant researcher at Harvard Medical School, Boston, who has called for a moratorium on trials.

But André La Prairie, an official at Health Canada, the country's equivalent of the FDA, says that "the guidelines are a clear message that limited controlled trials in xenotransplantation will continue". ■

... and sets up a body to oversee trials

Paul Smaglik, Washington

The US Department of Health and Human Services is to augment the panoply of government regulatory bodies by setting up a Secretary's Advisory Committee on Xenotransplantation (SACX) to oversee the technology.

The new committee will review proposed clinical trials and monitor ongoing trials, much as the Recombinant DNA Advisory Committee (RAC) of the National Institutes of Health (NIH) endorses novel gene-therapy protocols and monitors 'adverse events' in trials already under way.

But the role of the SACX in deciding which trials go ahead remains unclear. "There's been no decision yet on how it will work in terms of reviewing protocols," says Mary Groesch, a policy officer at the NIH's Office of Biotechnology Activities who is handling nominations for the SACX.

The Food and Drug Administration (FDA) will have the final say in approving clinical trials. In gene therapy, it is the NIH director, and ultimately the FDA, which has the last word.

In the early 1980s, the RAC approved the first human clinical

gene-therapy trial. By the 1990s, its responsibility was restricted to approving new types of protocol (see *Nature* 384, 297; 1996).

More recently, some critics have called for the RAC to stop examining new protocols altogether to focus on education, as the oversight of gene therapy by both the FDA and NIH has led to confusion, especially over reporting adverse events.

Researchers hope that, by sticking to a narrow mission, the SACX will attract less critical flak than the RAC. ■

US and Europe set up risk-assessment panel for GM crops

Munich US President Bill Clinton and Romano Prodi, the president of the European Commission, last week agreed to set up a panel of US and European scientists to address the possible environmental and health risks posed by genetically modified (GM) crops and other organisms.

The plan, announced at a summit in Lisbon, comes amid growing concern about a trade war between the United States and the European Union (EU) over GM products. US companies are concerned that European distrust of GM food could lead to import restrictions.

An EU spokesperson says that the panel's first priority will be to develop risk-assessment methods that all partners accept. A standard method is essential, she says, to provide hard scientific evidence of any potential risks. Details about the composition of the panel, and when it is expected to begin work, have yet to be announced.

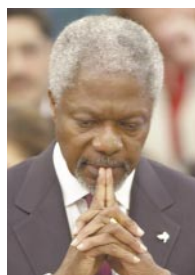
US laser report will not be ready until September

Washington The US Department of Energy last week missed a 1 June deadline to furnish Congress with a full prediction of the cost and schedule for the National Ignition Facility (NIF), the troubled laser project under construction at the Lawrence Livermore National Laboratory in California (see *Nature* 401, 101; 1999). In a memo sent to Congress on the day that the response was due, General Tom Giaconda, acting deputy administrator for defence programmes, said that a final cost and schedule will not be ready until mid-September.

An appropriations bill passed last year had also told the department to provide costs for closing down the project, but Giaconda declined to do so. The US administration is preparing a budget amendment that will provide an additional \$95 million to support NIF construction during 2001, with two-thirds of the money being diverted from other work at Lawrence Livermore and the rest from nuclear weapons research at other laboratories. It says its best current estimate is that the project will cost \$2.1 billion to build, compared with the original estimate of \$1.2 billion.

UN secretary-general puts faith in science

New York Kofi Annan, the secretary-general of the United Nations (UN), this week issued a statement emphasizing the need for "sound scientific information" to address the world's



Annan: seeking a lead from science.

deeply embedded in the fabric of our daily lives", and that "we are failing to invest enough in alternative technologies". Two weeks ago, Annan gave his support for a new initiative, the InterAcademy Council, designed to offer scientific advice to bodies such as the UN (see *Nature* 405, 268; 2000).

France invests in electronic encryption technology

Paris The French ministry of research last week launched a national programme to boost research in electronic encryption. The FF7 million (US\$1 million) programme will be led by Jacques Stern, director of the department of computer science at the Ecole Normale Supérieure in Paris, and a former head of the French computing firm Bull.

The programme will be advised by a 12-member commission drawn from the country's national research agencies, industry — including the telecommunications giant France Telecom — and a representative of the ministry of foreign affairs. The first 12 projects will favour young scientists, and research that involves the training of French scientists abroad — a call for proposals will go out this month.

Researcher sued over potential cancer drug

Washington Abbott Laboratories is suing cancer researcher Judah Folkman of Children's Hospital in Boston, over the intellectual property rights to part of a potential angiogenesis inhibitor. Angiogenesis inhibitors are considered valuable to cancer researchers and pharmaceutical companies because they could cut off a tumour's blood supply, thus starving it.

Folkman and his group have attracted considerable media attention during the past few years over two such putative inhibitors — endostatin and angiostatin (see *Nature* 393, 104; 1998). The dispute centres on who can claim credit for a portion of a protein, kringle 5, that may have angiogenesis-inhibiting properties.

In a complaint filed in US District Court in Boston, Abbott claims that Folkman "fraudulently" filed a patent claim on kringle 5. Folkman told the *Boston Globe*

environmental problems. Such information, he said, was "the only basis for effective policy," but he admitted that "large gaps in our knowledge remain".

In his statement, issued on World Environment Day, Annan said that "unsustainable practices remain

that: "The dispute comes down to Children's Hospital wanting to protect its patent rights and Abbott wanting to avoid paying any royalties."

New Zealand climbs aboard South African telescope

Sydney New Zealand astronomers, frustrated by their limited viewing facilities — one small observatory at the University of Canterbury with a modest 1-m and two 60-cm telescopes — are celebrating their acceptance last week into the Southern African Large Telescope (SALT) project.

SALT is building a 10-m telescope, the largest in the Southern Hemisphere, with German, Polish and US funding (see *Nature* 402, 452; 1999). New Zealand will now contribute US\$1.3 million in cash and kind. The latter will be the design and building of a high-resolution spectrograph, a field in which the country's physicists excel.

Compton falls to Earth out of harm's way

Washington The US space agency NASA has brought the Compton Gamma-Ray Observatory (CGRO) safely down into the Pacific Ocean. Safety concerns had led NASA to de-orbit CGRO in a remote region as, if a second of its three gyroscopes had failed, the satellite's uncontrolled re-entry would have posed a 1/1,000 risk of loss of human life.

Weighing around 15,000 kg, the observatory was one of the heaviest spacecraft ever launched, and pieces as large as one tonne were expected to survive re-entry. The observatory had exceeded its mission period by four years and made an important contribution to our knowledge and understanding of gamma-ray astronomy (see *Nature* 405, 504–506; 2000).

Venter and Wilson give prize money to charity

Washington Craig Venter, president of Celera Genomics in Rockville, Maryland, has announced that he will donate his portion of the 2000 King Faisal International Prize of Science to The Institute of Genomic Research (TIGR), the non-profit research institution he founded. TIGR will put the \$100,000 toward sequencing the genome of *Theileria parva*, a protozoon transmitted by ticks that causes a leukaemia-like disease in cattle.

Venter split the \$200,000 prize with Edward Wilson, honorary curator in entomology of Harvard University's Museum of Comparative Zoology and Pulitzer-prizewinning co-author of *The Ants* (Springer, 1990). Wilson is giving his portion of the prize to the E. O. Wilson Foundation, which will support biodiversity research.

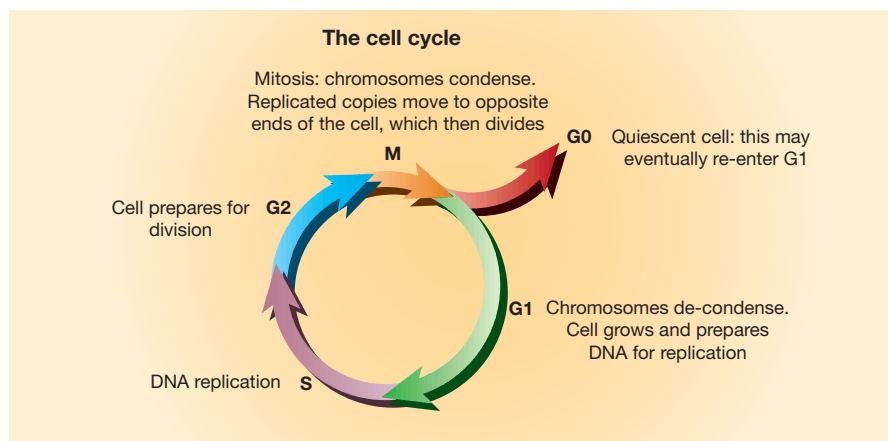
Cloning's owners go to war

The team that created Dolly the sheep captured the headlines, but several groups now have patents on cloning. Peter Aldhous considers how this tangled web of proprietary claims will affect the future of the technology.

To the uninitiated, it seemed like an arcane scientific argument. In September 1998, some 18 months after Dolly the cloned sheep was unveiled to the world, her creators issued a challenge. They disputed one of the conclusions drawn by a group claiming to have repeated the feat of cloning animals, this time in cattle.

Ian Wilmut, of the Roslin Institute near Edinburgh, and Keith Campbell, then with a Roslin spin-off company, PPL Therapeutics, did not question that central claim. But in a letter to *Science*¹, they took issue with their rivals' assertion that the donor cells used to clone the calves had been actively dividing. The cow cloners, from the University of Massachusetts in Amherst, and Advanced Cell Technology (ACT), a company in nearby Worcester, defended their interpretation².

This was no ordinary scientific spat. The Roslin researchers' challenge is central to the relative value of the patents that they, and their rivals, have subsequently been awarded. And with others claiming rights to further pieces of intellectual property in cloning, the battle lines are being drawn. Already, one cloning patent infringement suit has come to court, and with more groups muscling in with their own claims, it seems unlikely to be the last. "It's a very tangled web," says Steven Stice, formerly with



Taking a breather: many cells opt out of the cell cycle of growth and division to sit in a quiescent state.

the Massachusetts team and now at the University of Georgia in Athens.

The technique used to make Dolly is called nuclear transfer cloning. The idea is to take an egg cell, strip out its chromosomes, and then fuse it with another cell, which donates its nucleus to the egg. If the egg is then activated so that it starts dividing, it can develop into an animal that is a clone of the individual from which the donor cell was taken.

Scientists interested in animal production had been experimenting with nuclear

transfer for many years. But they had only succeeded using undifferentiated donor cells from very young embryos. The consensus emerged that, in mammals, it was impossible to produce clones from differentiated cells, specialized for a particular function.

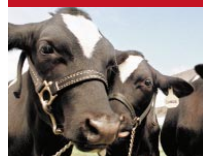
Quiet life

Dolly, cloned from a cell derived from an adult ewe³, conclusively overturned that dogma. But in truth, the real breakthrough had come in March 1996, when the Roslin team described two lambs, Megan and Morag, cloned from cells that came from fetuses, but were nonetheless differentiated⁴.

The Roslin researchers thought the key to such cloning was sending the donor cells into a quiescent state known as G0. Cells that are proliferating go repeatedly through the series of phases in the cell cycle — growing, copying their DNA, and dividing (see above). But many cells cease dividing and sit in the G0 state. In laboratory cultures, cells can be forced into the G0 state by depriving them of nutrient-rich serum. And that is what Campbell, the Roslin team's cell biologist, did. In August 1995, the Roslin researchers filed patents claiming rights to cloning from differentiated, quiescent cells.

At face value, the Massachusetts researchers were way off the pace. In May 1998, they published a paper that repeated the Megan and Morag accomplishment with cows⁵. But in the context of intellectual property, it was significant. The researchers, led by Stice and his colleague James Robl, took their donor fetal cells from an actively dividing culture, missing out

Making cloning pay



PRODUCING ANIMALS OF HIGH GENETIC VALUE

If you have managed to breed a cow that regularly breaks records for its milk yields, why risk losing its

combination of high-quality genes in the lottery of sexual reproduction? Cloning offers a means to create herds of carbon-copy animals.

PHARMING HUMAN PROTEINS

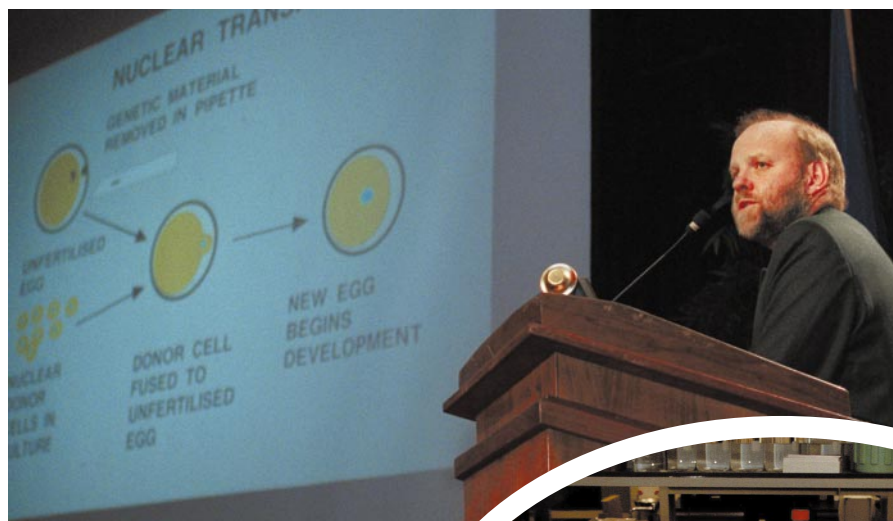
Many human diseases — haemophilia, for instance — are caused by defects in the production of proteins. Several companies are now trying to make genetically engineered animals that secrete human proteins in their milk. By combining cloning with a transgenic technology called gene targeting, herds of these 'pharm' animals can be created.

XENOTRANSPLANTATION

Pigs could provide a supply of organs for patients on transplant waiting lists, if the organs can be genetically engineered so that they do not immediately trigger rejection by the human immune system. Again, cloning can be combined with gene targeting to produce these precious animals.

THERAPEUTIC CLONING

Here the goal is to repair the human body with cell and tissue grafts that are perfectly matched to the recipient. Just take a healthy cell from the patient, use it to create a cloned embryo, then — after just a few days — dissect it to remove the embryonic stem cells that, in theory, can be grown in culture to give rise to any of the body's tissues. Working out how to achieve that, however, poses huge scientific challenges.



Property dispute: both Wilmut (above) and the team at Advanced Cell Technology (left to right: West, Cibelli and Robl) hope their patents on cloning will dominate the field.

the serum starvation step. They described these cells as “non-quiescent”, and argued that they were in a phase called G1, in which cells grow and prepare to copy their DNA. More than a year before, in January 1997, they had applied for a patent. It covered cloning from proliferating differentiated non-human cells — that is, cells in any phase except G0.

As cloning has a range of possible commercial applications (see ‘Making cloning pay’, opposite), the Roslin team had to respond. The group argued that the Massachusetts researchers could not prove that the individual cells from which they had cloned their calves were not in G0, even though the cells were in a dividing culture — a view that Wilmut still holds. “To my knowledge, there is nothing yet published that clarifies the situation,” he said, when questioned by *Nature* on the issue after a lecture in London last month.

Any cell will do?

It is difficult to determine precisely where in the cell cycle an individual cell is. But now that many groups have experimented with cloning from proliferating cell cultures, it is hard to find anyone who thinks that putting donor cells into G0 is the key. “It is probably going to be shown that quiescence is not an essential factor,” says Davor Solter of the Max Planck Institute for Immunobiology in Freiburg, Germany. Even Campbell, now at the University of Nottingham, acknowledges the consensus. “I’ve always been very careful to say that G0 may be beneficial,” he says. “It doesn’t mean that other stages might not work.”

Meanwhile, the disagreement has shaped up into a tussle between ACT and a rival company, Geron of Menlo Park in California. The US patent granted to the University of Massachusetts in August 1999 is exclusively



licensed to ACT. Geron won extensive rights to the British patents that were eventually awarded to Roslin in January this year, by acquiring a spin-off, Roslin BioMed, in May 1999.

ACT’s chief executive officer, Michael West, knows Geron well — he was one of its founders. “They’ve got a piece of the pie, and we’ve got the rest of the pieces,” he says. Not surprisingly, Geron sees it differently. David Earp, the company’s vice president for intellectual property, argues that it should be possible under US patent law to extend the claims of the Roslin patents to all forms of cloning using adult cells. “Our intellectual property estate eventually will acknowledge the historic breakthrough represented by the cloning of Dolly,” he asserts.

Maybe so, but many observers believe ACT has played a clever game. Alan Colman, research director at PPL Therapeutics, describes the broad claims in its patent as “opportunistic”, and says he was surprised they were granted. “But it will be quite difficult to get rid of the ACT patent,” he adds.

But it is not just the basic patents on cloning from differentiated cells that will determine who makes a commercial success of the technology. To bring many applications to market, nuclear transfer will have to be combined with other technologies. To clone transgenic animals that produce valuable human proteins in their milk, for instance, it must be combined with techniques for creating these animals. For these applications, Roslin has licensed its

patents to PPL Therapeutics, which is developing its own transgenic technology. ACT, meanwhile, has linked up with Genzyme, a company based in Cambridge, Massachusetts.

Geron is interested in therapeutic cloning, which aims to grow human tissues to replace those that are damaged or diseased. The company has exclusive access to two other technologies that it sees as crucial. First, it holds licenses to patents on human embryonic stem cells, which would be isolated from cloned embryos a few days after their creation, and from which replacement cells and tissues could be grown. Second, it controls patents on telomerase, an enzyme that rebuilds the caps,

or telomeres, on the ends of chromosomes. These caps shorten as cells age, so to ensure that cloned tissue grafts are vibrant and healthy, Geron says, it will be necessary to rejuvenate the donor cells using telomerase.

But true to form, ACT claims it can stake out its own position. The company’s existing cloning patent is not relevant here, as it only covers nuclear transfer from a non-human donor cell. But, controversially, ACT revealed in November 1998 that one of its scientists, Jose Cibelli, had fused cells taken from his body with bovine eggs, stripped of their chromosomes, to create embryos which were allowed to grow for several days. Arguably, this would allow ACT to generate embryonic stem cells that would not be covered by Geron’s patents, as they could not be regarded as completely human.

Cibelli’s experiments are the subject of a patent application from ACT. And in April this year, ACT suggested that the act of cloning itself rejuvenated donor cells in cattle, rebuilding their telomeres to beyond the length seen in newborn calves⁶. “We believe we can make therapeutic cloning work without transgressing any of the Geron patents,” says West.

Animal pharm

The idea of using human–cow hybrid cells for therapeutic cloning is viewed with scepticism by many experts — even if it works, the regulatory hurdles would be immense. What is more, Cibelli’s work has never been published in a peer-reviewed journal. But biologists are intrigued by ACT’s telomere paper⁶, and the implication that therapeutic cloning need not depend on telomerase. “It might be it’s necessary, but I don’t think we have enough information to say,” says Colman.

Other companies are adding to the confusion. Most significantly, ACT has found itself on the wrong end of a patent infringement suit from Infigen, a company in DeForest, Wisconsin. Infigen has long been interested in cloning

cows of 'high genetic value', and has recently teamed up with the American Red Cross and Pharming, of Leiden, the Netherlands, to clone animals that produce human proteins in their milk. The company is also working with Imutran of Cambridge, England, and the Swiss-based drugs giant Novartis, on cloning in xenotransplantation — the use of organs from animals in human transplantation.

Staking a claim

Infigen owns a suite of patents on the basic techniques of nuclear transfer, awarded before the Roslin team demonstrated that it is possible to clone mammals from differentiated cells. Last year, it sued ACT for breaching two US patents on cow cloning, one covering a specific culture medium, the other a method for activating bovine eggs after transferring the donor nucleus. Infigen also claimed that Stice, who had once worked for Infigen, stole its trade secrets. That complaint was rejected, but in June 1999, the US District Court in Wisconsin ruled that ACT had indeed infringed Infigen's patents — after which the two companies came to a confidential settlement.

Ominously, it seems that other cow cloners could soon be hearing from Infigen's lawyers. "We're taking steps right now to inform several parties about our patent estate," says Michael Bishop, the company's vice-president for research.

Infigen gained a further US patent in January this year, which is causing raised eyebrows. Again specific to cows, this patent covers cloning from fetal cells that have been developmentally 'reprogrammed' by treating them with specific biochemical growth factors. Given that the Roslin team and others have shown that nuclear transfer itself can reprogramme differentiated

Paradise lost in Hawaii

No discussion of cloning's landmark achievements would be complete without mentioning 'team Yana'. Working in the lab of Ryuzo Yanagimachi at the University of Hawaii in Honolulu, scientists led by Teruhiko Wakayama (below) stunned the world in 1998 by cloning scores of mice, some of them clones of clones⁷.

Rather than fusing a donor cell with an egg, like other groups working in the field, Wakayama developed a technique in which he removed the donor cell's nucleus and injected it directly into the egg using a piezoelectric device. The university filed for a patent on the method, and granted an exclusive license to a local company called ProBio, headed by

Australian businessman, Laith Reynolds.

But the story has since gone sour. ProBio is still waiting for the cloning patent to be granted, but 'team Yana' has broken up. Tony Perry (right), a member of the team, is suing the university over the rights to a transgenic technology, now licensed to ProBio, which was developed by him while he was a European Molecular Biology Organization research fellow in Yanagimachi's lab. Although Wakayama has not sued over the rights to his cloning technique, he is understood to be similarly unhappy. Both scientists have now left for Rockefeller University in New York, but declined to discuss the reasons for their move with *Nature*.



cells, most researchers cannot understand the relevance of the extra step.

Although the significance of Infigen's new patent remains unclear, everyone in the field is watching for the emergence of other patents that could alter the intellectual property landscape. "In my mind, there's no over-arching patent out there," says Stice. "We are still all trying to find one technique that is efficient." Indeed, most cloning groups only get one or two live births for every hundred nuclear transfer procedures they perform.

Many observers are keeping a close watch on PPL Therapeutics, which in March announced that it had cloned five pigs from adult cells using a novel technique. Details have not yet been published, but

Colman

claims the method is "significantly different" from anything described previously. That could be important, as pigs have proved hard to clone, and are the animals of choice for xenotransplantation. An Australian company, Stem Cell Sciences of Melbourne, also claims to have developed an alternative method of cloning that similarly remains under wraps.

"The intellectual property situation in this field is very complex at the moment," says Emma O'Donovan, editorial analyst at Derwent Information, a company in London specializing in patent information. "The wording and scope of individual claims will have to be examined very carefully."

If the current confusion is not resolved, the danger is that the situation will restrict the flow of money needed to develop the technology. Investors like the ownership of the key intellectual property to be clear, explains Neal First of the University of Wisconsin in Madison, a cloning specialist who sits on the board of a venture capital company interested in the field.

But perversely, the confusion could have a stimulating effect in the short term. "There are scientists starting little companies without much regard to where the intellectual property lies," says First. "And they are increasing the pool of knowledge." But if the writs begin to fly, some of these scientists may wish they had been more circumspect.

Peter Aldhous is Nature's Chief News and Features Editor.

1. Wilmut, I. & Campbell, K. H. S. *Science* **281**, 1611 (1998).
2. Robl, J. M., Jerry, D. J., Stice, S. & Cibelli, J. *Science* **281**, 1611 (1998).
3. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385**, 810–813 (1997).
4. Campbell, K. H. S., McWhir, J., Ritchie, W. A. & Wilmut, I. *Nature* **380**, 64–66 (1996).
5. Cibelli, J. B. *et al. Science* **280**, 1256–1258 (1998).
6. Lanza, R. P. *et al. Science* **288**, 665–669 (2000).
7. Wakayama, T., Perry, A. C. F., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. *Nature* **394**, 369–374 (1998).



Quintuple vision: PPL Therapeutics has devised a new technique that has enabled it to clone five piglets.

Regulation, not private enterprise, is the key to a healthy environment

Sir—Gretchen C. Daily and Brian H. Walker write in their Commentary “Seeking the great transition” (*Nature* **403**, 243–245; 2000) about the importance of involving the private sector in maintaining a stable and wholesome environment. They are, of course, correct, and industry has made a variety of efforts towards improving efficiency, recycling waste and conserving environmental quality.

However, many other examples reveal that the private sector cannot be left to regulate itself. In general, the purpose of industry is profits, not the public good or public service. Protracted fights over lead in petrol, agricultural poisons in food, and industrial waste contaminating rivers and land have burned into the public’s mind the necessity for strict regulation of industrial practice. Blatant distortions of truth by the tobacco industry and the Global Climate Coalition have added to this concern. Simple exhortations are not enough.

I agree with Daily and Walker that the scientific community must participate more in defining where the public interest lies. But this definition has to do with the chemistry and physics of environment and the rules necessary for keeping the environment functioning as an appropriate place for humans and other life. Economic and political objectives are inadequate, except as they strengthen compliance with ecological and environmental needs.

The public must keep asking questions: what is clean water, clean air, an appropriate place to live, and appropriate behaviour on the part of one’s close neighbours?

We look to governments to establish those rules. Where central governments fail, the rules are established locally, as for example in many communities in the Amazon Basin. The realization of the need for rules emerges from the recognition that resources are limited and are contested among diverse interests — in many cases including exogenous industry seeking an opportunity to exploit resources already being used by indigenous people.

Democratic capitalism must be developed much more intensively and deliberately. Scientists can play their part, by defining what is required for a human habitat so that the public can know what is in its interest and what is not. Relinquishing those decisions to commerce and industry has never worked in the past, it does not work now, and it will not work in the future.

Hence, although I agree with much of Daily and Walker’s argument, I think they are too dismissive of governments’

potential to address the challenges of protecting our environment.

George M. Woodwell

*The Woods Hole Research Center, PO Box 296,
Woods Hole, Massachusetts 02543, USA*

England and US corner the journal market

Sir—Here I describe what I believe is the first study of the distribution of biomedical publications in Medline with respect to publishers’ location. Previous studies have looked at the location of the first author, including the number of publications by country¹, by US state² or in the European Union³, normalized to population size, number of doctors in the country concerned⁴ or gross domestic product⁵.

I searched every year from 1966 to 1998, using Medline Express, selecting the field “country of publication” combined with “publication year”. I carried out this search for 194 countries. The number of publications listed in Medline increased every year: from 174,584 in 1966 to 417,944 in 1998, an average annual increase of 2.76%.

Just 12 countries (Fig. 1) each published more than 1% of the total: during 1996–98, their combined share averaged 88.9%. Publishers from 25 other countries contributed 0.1–1%; 28 countries 0.01–0.1%; and 34 countries less than 0.01% (but more than zero).

Surprisingly, 95 countries failed to contribute a single article. Most of these are smaller, underdeveloped countries with low scientific output, possibly reflecting Medline’s exclusion of journals published in developing countries⁵.

When I analysed the 12 countries in which most journals are published, four had increased their share of publications: the United States was up from 31.64% in 1966 to 46.96% in 1998; England from 9.32 to 18.46%; the Netherlands from 2.01 to

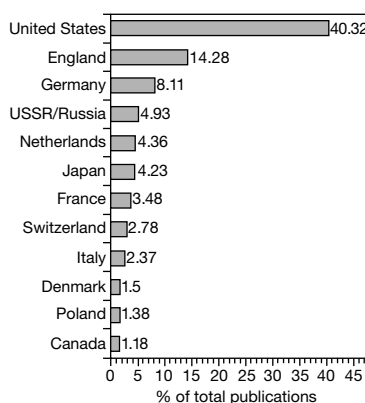


Figure 1 The percentage of biomedical publications in journals edited in the 12 most prolific countries from 1966 to 1998.

5.26%; and Denmark from 1.15 to 1.65%. Publishers from the 12 countries contributed 90.1% of the total output in 1998, compared with 86.19% in 1966.

The proportion decreased in Germany from 9.85% to 5.79%; the Soviet Union/Russia from 8.67 to 1.52%; Japan from 5.08 to 3.08%; France from 5.92 to 2.14%; Switzerland from 2.6 to 2.19%; Italy from 5.54 to 1.33%; Poland from 2.87 to 0.59%; and Canada from 1.54 to 1.13%.

A. A. Waheed

*Department of Protein Biochemistry, Tokyo
Metropolitan Institute of Gerontology, 35-2 Sakae-
Cho, Itabashi-ku, Tokyo 173-0015, Japan*

1. Benzer, A. *et al. Lancet* **341**, 247 (1993).
2. Thompson, D. F. N. *Engl. J. Med.* **340**, 817–818 (1999).
3. Hefler, L., Tempfer, C. & Kainz, C. *Lancet* **353**, 1856 (1999).
4. Eldor, J. *Lancet* **341**, 634 (1993).
5. Rosselli, D. *Lancet* **354**, 517 (1999).

Celera’s role in opening up new frontiers

Sir—I would like to set the record straight about my closing statement at the House Science Subcommittee on Energy and Environment hearing on the Human Genome Project (*Nature* **404**, 691; 2000).

Your reporter wrote that I said “Government’s role is to initiate research on which private sector companies can capitalize, pointing to Celera Genomics as a prime example” and that I “endorsed the way that Celera is incorporating the public Human Genome Project’s data into its own database, adding its own sequence information, providing annotation tools, then selling it”.

This is a misleading interpretation of my views, and of comments that I made to your reporter and during the hearing (see http://www.house.gov/science/106_hearing.htm#energy_and_environment). What I sought to express is an optimistic view of the benefits of federal research, including technology transfer to the private sector.

In closing, I noted the role of the US federal government in settling the American West and in the development of the Internet. I said “Think what would have happened if the Defense Advanced Research Projects Agency had stepped forward and insisted that the public interest could not be served unless it controlled the content and rate of growth of the Internet. Dr Venter and others are responsible for speeding up the sequencing of the human genome by five years. For this reason at least, I would rather have the problems of private-sector involvement in the human genome field than not. Some problems are good to have, and I think this is one of them.”

Ken Calvert

*US House of Representatives Subcommittee on
Energy and Environment, Rayburn House Office
Building, Washington DC 20515-6301, USA*

Climate change in perspective

Our concerns about global warming have an age-old resonance.

Hans von Storch and Nico Stehr

Many people think that the threat of 'global warming' arose only towards the end of the twentieth century. But the idea of human (anthropogenic) interference with climate has an important — although often overlooked — historical dimension. Climate change, either natural or anthropogenic, has been discussed from the classical age onwards, evolving from the expected benefits of climate engineering to today's fear of global disaster.

People have always been aware of climatic variation — in the freezing of rivers or success of harvests, for instance¹. Whether these changes were interpreted as natural or man-made depended mainly on the philosophy of the time. In the Middle Ages, change was seen as a natural process, the systematic deterioration of a living, and thus ageing, world. Anthropogenic changes were thought of as man's attempt to fulfil God's task of completing creation, whereas negative events were seen as a heavenly riposte to bad behaviour, such as punishment for tolerance towards witches².

By the eighteenth and nineteenth centuries, interest focused on the effects of deforestation and other forms of land use. The colonies of North America, for example, were thought to have become more temperate as a result of deforestation during colonization³. By 1890, climate change was discussed in modern terms: "Deforestation, as a part of agricultural expansion everywhere, must necessarily result in less rainfall and more frequent droughts. This view is most poignantly expressed by the saying: man walks the earth and desert follows his steps! ... The Italian government has been paying special attention to reforestation and its expected improvement of the climate ... The alternation of periods of heavy rainfall with droughts must be prevented ... In the United States, deforestation is seen as the cause of a reduction in rainfall ... The committee chairman of the AAAS has demanded decisive steps to extend woodland to counteract the increasing drought. In Vienna in 1873, the Congress for Agriculture and Forestry discussed the problem in detail. When the Prussian house of representatives ordered a special commission to examine a proposed law concerning the preservation and planting of forests, it pointed out that the falling water levels of Prussian rivers was one of the most serious consequences of deforestation, and could only be reversed by reforestation. The same concerns were raised in Russia."^{4,5}

Optimism about the benefits of climate engineering was replaced with fear of 'climate weapons'.

More exotic ideas about human influence on climate mirrored technological developments. Lightning conductors (in 1816 in Switzerland), nuclear explosions in the 1950s, supersonic transport in the 1970s, and space traffic have all been blamed for climatic deterioration⁶. In the first half of the twentieth century, optimism about the potential of science and technology was reflected in plans to improve climate on an almost hemispheric scale by redirecting the Gulf Stream or Siberian rivers, or blocking the Bering Strait. Later, this optimism was replaced by concern that adversaries might develop 'climate weapons' — perhaps for modifying the global ocean circulation — and an international agreement banning them was prepared⁷. The possibility of a nuclear winter following from military action was discussed in the 1980s, as was the effect of burning oil wells in Kuwait at the beginning of the 1990s.

The modern concern about carbon dioxide as an agent of anthropogenic climate change can be traced back to the end of the nineteenth century and the Swedish chemist Svante Arrhenius. In 1933, a paper in *Monthly Weather Review* identified a significant warming trend, which in 1938 was related to the human greenhouse effect. This warming was followed by a global cooling, thought to be the first indication of a new ice age, accelerated by aerosols from industrial pollution blocking out sunlight. Eventually, Arrhenius's theory was revived, supported by the strong warming in the 1980s and 1990s, palaeoclimate findings and sophisticated modelling studies. As in the nineteenth century, the concept of anthropogenic climate change became news, advocated by scientists. Government bodies were established to advise on suitable action.

Although most of the historical concerns over climate turned out to be exaggerated, we are not claiming that the present concept of global warming is flawed. We are convinced that greenhouse gases are accumulating in the air, and strongly believe that near-surface temperatures are rising in response. But we

are not convinced that present and future climate change will have a significant impact on society and global ecosystems.

Today, as in the past, the claim that anthropogenic climate change is associated with a serious impact on human society is still a hypothesis, often based on simplistic methodology. An important contrast with many of the earlier cases is that climate change is now perceived as negative. Of course, this view is not limited to climate change. Until the 1950s, it was thought that science would improve living conditions. Nowadays, it is often seen as a threat. When modern science scribbles on the wall, it is no longer about emancipation from nature but about possible disaster — nuclear war, genetic manipulation, climate change.

Many scientists realize that our knowledge of the climate system will always suffer from significant uncertainty because of its open, complex and heterogeneous character and the long timescales involved. Thus, studies of climate change are bound to be characterized by high uncertainty and high stakes, with public, antagonistic debates not only between scientists but also activists and other non-specialists^{8,9}. However, age-old concerns about extremes of climate are part of the cultural background², for scientists as well as the public. This subjectively genuine concern, nourished by the mass media, underlies the activist scientists, who wield the same apocalyptic scenario of drought, delugial floods and devastating storms as their historical counterparts. ■

Hans von Storch is at the Institut für Gewässerphysik, GKSS Forschungszentrum, Max-Planck-Strasse 1, D-21502 Geesthacht, Germany.

Nico Stehr is at the Sustainable Development Research Institute, University of British Columbia, 6201 Cecil Green Park Road, Vancouver V6T 1Z1, Canada, and the Fachbereich Soziologie, Universität Duisburg, D-47057 Duisburg, Germany.

1. Lamb, H. H. *Climate, History and the Modern World* (Methuen, London, 1982).
2. Glacken, C. J. *Traces on the Rhodian Shore* (Univ. California Press, Berkeley, 1967).
3. Williamson, H. *Trans Am. Phil. Soc.* 1 (1771).
4. Brückner, E. *Klimaschwankungen seit 1700 nebst Bemerkungen über die Klimaschwankungen der Diluvialzeit: Geographische Abhandlungen herausgegeben von Prof. Dr. Albrecht Penck in Wien* (Holzel, Vienna and Olmütz, 1890).
5. Stehr, N. & von Storch, H. (eds) *The Sources and Consequences of Climate Change and Climate Variability in Historical Times* (Kluwer, Dordrecht, 2000).
6. Kempton, W., Boster, J. S. & Hartley, J. A. *Environmental Values in American Culture* (MIT Press, Cambridge, MA, 1995).
7. Ponte, L. *The Cooling* (Prentice-Hall, Englewood Cliffs, 1976).
8. Functowicz, S. O. & Ravetz, J. R. in *Risk Analysis in the Private Sector* (eds Whipple, C. & Covelio, V. T.) 217–231 (Plenum, New York, 1985).
9. Fleming, J. R. *Historical Perspectives on Climate Change* (Oxford Univ. Press, 1998).

Skull and cross words

How the discovery of an ancient American skeleton set science against law.

Riddle of the Bones: Politics, Science, Race, and the Story of Kennewick Man

by Roger Downey

Copernicus: 2000. 216 pp. \$25, £17

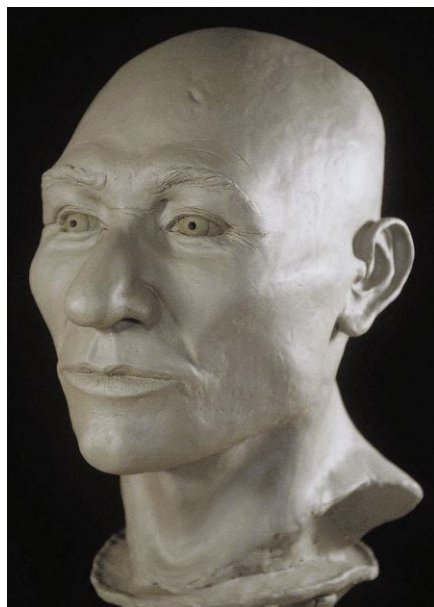
Simon Conway Morris

In his famous work *Democracy in America*, first published in 1835, the French political scientist Alexis de Tocqueville wrote, "I thought that the English constituted the most serious nation on the face of the Earth, but I have since seen the Americans and have changed my mind". Nothing much seems to have changed in the intervening 160 years, at least to judge from this book, which documents with painful clarity the continuing circus that passes, at least in some quarters, for American life. To the indigenes no doubt it is all terribly important, but, viewed remotely, a rather different image swims into view: the all-too-familiar story of vaunting ambition, mad bureaucracy, political correctness and far, far too many lawyers, all snugly cemented by the glue of mutual suspicion and distrust.

Kennewick Man joins that long succession of archaeological finds that have had imposed on them cultural burdens and assumptions that far outstrip any legitimate reading of the evidence, which as often as not is thin in the extreme. The discovery of a partial skeleton by two teenagers wading in the shallows of the Columbia river was swiftly followed by its appropriation by a local archaeologist, one Jim Chatters.

Like various other characters that litter the pages of *Riddle of the Bones*, Chatters is not described in exactly glowing terms by Roger Downey. But he is first on the scene and gives the juggernaut its initial push. Radiocarbon dating soon proves Kennewick Man to be satisfactorily old, and there is intriguing evidence of a stone spear-point embedded in the pelvis. The sensation, however, lies in Chatters' examination of the skull. What emerges, as Chatters makes a series of frantic notes, is not an ancient or palaeo-Indian, a poignant relic of the last great human diaspora that led to the overrunning of the Americas. Instead, against all expectation, features of the skull appear, incredibly, to be 'Caucasoid', that is, most similar to a 'pre-modern European'.

Now the juggernaut begins to pick up speed, and the first hurdle seems to be easily surmounted. The skeleton was found on federal land, under the jurisdiction of the US Army Corps of Engineers. Their hallmark is one of unrestrained hydraulic enthusiasm, with a principal contribution to the Ameri-



Origins of Kennewick Man: his unexpected 'Caucasoid' features made him a coveted prize.

can landscape being the erection of innumerable dams, irrespective of the mounting evidence for resultant environmental mayhem. A force to be reckoned with, but in any event Chatters' conveniently postdated excavation permit is swiftly obtained.

Things begin to unravel, however, at the next obstacle, which arises in the form of NAGPRA, the innocuous acronym for the 1990 Native American Graves Protection and Repatriation Act. Chatters' dispatching of a bone for radiocarbon dating, and thereby at least the partial destruction of the skeleton, was, as he should have known, a potentially serious mistake. Set up after years of effectively unrestricted archaeological excavation of Indian sites — looting, if you prefer — and the amassing of skeletons and artefacts in museums across the United States, NAGPRA was a well-meant but muddled and inconsistent piece of legislation designed to restore some semblance of dignity to the religious and cultural sensitivities of North America's original inhabitants.

The catch, of course, is that, for all intents and purposes, much of Indian history will remain forever irrecoverable unless it is subjected to scientific analysis. Almost inevitably, Kennewick Man, 'Caucasoid' or not, was now inextricably caught up in the turmoil of archaeological and museum politics, legal battles, Indian rights and the elephantine process of federal bureaucracy. Few of the characters depicted emerge with much credit as e-mails and telephones hum with acerbic demands and face-saving

manoeuvres, steadily enmeshing an increasingly wide circle of individuals.

The cast is large, and Downey takes the narrative at a spanking pace — larded with folksy asides — that leaves the reader breathless to discover the final denouement. It slowly becomes clear, however, that, despite its tragicomic tone, the real target of *Riddle of the Bones* is not the various twists and turns of the American legal system, nor the Odin-worshipping Asatru Folk Assembly, who greet Kennewick Man as a useful prop for their white supremacist world-view, or even the genuine ambiguities of Indian ownership, physical or spiritual, of archaeological remains that conceivably in reality derive from some long-vanished tribe or nation. Rather, as I read this book, it is as much to do with our, that is the West's, increasingly ambiguous attitude towards 'Science' with a big S.

Downey's views are marred by some garbled sections, not least the supposition that mitochondria were once virus-like, but, despite a broad sweep of current concerns, the underlying tenor is an uneasy mixture of relativist belief that ultimately nothing can really be known and a misplaced mathematics-envy that presupposes that only terribly clever people can really understand science. Downey's confusion of values is clear enough when, in self-contradictory style, he writes of one method (analysing teeth) as looking "behind the firm façade of numbers, charts and diagrams", while a little later he laments that "statistical thinking ... is beyond the capacities of many anthropologists". But he then continues, "They must, like the rest of us, take such results on faith, trusting to others more deeply indoctrinated to raise an alarm should the methodology or results seem dubious". Façades, indoctrination?

But, for all that, Downey deserves neither our scorn nor our condescension, because it is only in the last few pages that his message suddenly crystallizes. We never really learn who or what Kennewick Man was, or still is, but in his conclusion, reached as he sits perched on a spot known as Jump Off Joe, Downey achieves a near-mystical vision of combined paradoxes — continuity and aloneness, faith and knowledge — that reminds us that, for all the antics associated with Kennewick Man, where we came from and why we matter are genuinely important questions that science alone is ill-equipped to answer.

Simon Conway Morris is in the Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK.

The quantum jungle revisited

The New World of Mr Tompkins

by George Gamow & Russell Stannard
Cambridge University Press: 1999. 256 pp.
£14.95, \$24.95

Anton Zeilinger

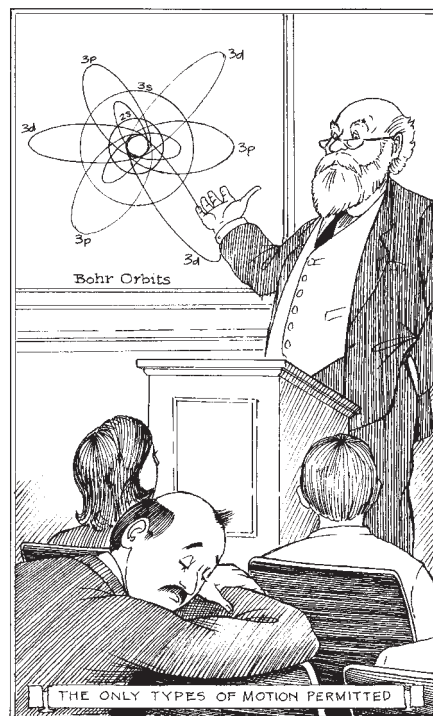
Have you ever wondered what the twentieth century's two great theories of physics — quantum theory and the theory of relativity — really tell us? Every physicist views these theories as providing a major break with the world-view — generally known as classical physics — that existed until the end of the nineteenth century.

The problem with considering these phenomena, and the reason the theories were not discovered earlier, is mainly one of scale — relativity theories are concerned with the enormously high speed of light, while quantum mechanics is characterized by Planck's extremely tiny quantum of action. Thus, in general, to directly 'see' relativistic phenomena one must go to speeds far beyond the scale people are familiar with, whereas quantum phenomena can only be seen by analysing changes occurring on extremely small scales.

In 1938, the Russian-born physicist George Gamow, who was at that time living in the United States, invented the elegant idea of considering what phenomena we would experience if the velocity of light was not 300,000 kilometres per second, but instead was the speed achieved by a fast cyclist, and if Planck's quantum of action was large enough for quantum phenomena to be part of everyday experience. Gamow invented "Mr Tompkins", the bank clerk who, curious about modern physics, attended public lectures at the local university. However, having a rather short attention span, he fell asleep during the lectures and dreamt of living in just such a strange world. In his dreams he rode a relativistically flattened bicycle, played billiards where quantum uncertainty added greatly to the excitement of the game, and met tigers that had been quantum-diffracted by the trees in the jungle.

Gamow's book was published in 1939 as *Mr Tompkins in Wonderland*, with a sequel, *Mr Tompkins Explores the Atom*, in 1944. The two books appeared together as *Mr Tompkins in Paperback* (Cambridge University Press) in 1965.

While these charming stories have been a source of great pleasure to many physicists, some are unfortunately becoming outdated, although they remain interesting historical documents. Given the developments of physics in the second half of the twentieth century, the physicist Russell Stannard, a popular-science writer perhaps best known for his *Uncle Albert* children's trilogy involving the adventures of Einstein, undertook the task of



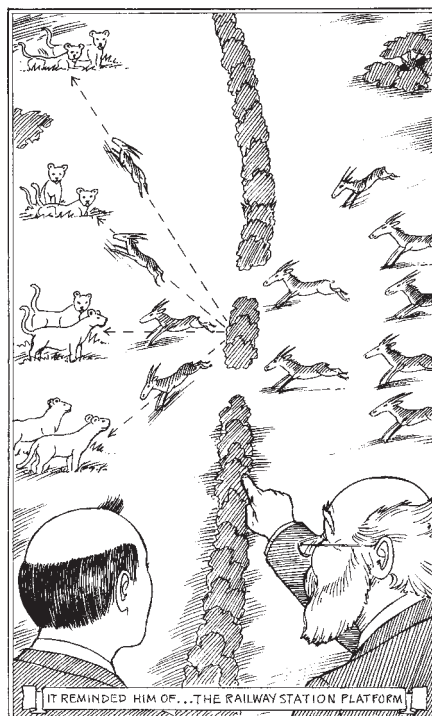
Quantum dreaming: Mr Tompkins' short attention span delivered him to a strange other world.

rewriting some of the stories and adding new material. Particularly useful are his additions on cosmology, which include black holes and the modern unified theory of particle physics. There is a beautiful explanation of the SU(3) representation, which explains how elementary particles can be understood on the basis of deeper symmetries, and how quarks are able to synthesize all the new particles observed in high-energy physics.

Certainly the most difficult task, both in Gamow's time and today, is to explain quantum mechanics to the general public. This is because its concepts are completely counterintuitive, and because even today's physicists disagree about what it means philosophically. So it is not surprising that Gamow's original presentation may not be totally satisfactory today.

Stannard has rewritten a significant part of Gamow's original chapters on quantum mechanics, but not always in a way that makes them clearer. Whereas Gamow began his explanation using an analogy to billiard balls, without explaining clearly why the concept of quantum mechanics is so strange and thus leaving the reader pretty much in the dark, Stannard starts by describing how we cannot obtain a precise measurement for the track of an elementary particle because of the unavoidable disturbance caused by the measuring process. Although the discussion is valid, it might leave the reader believing that, while we cannot precisely determine, say, the properties of an elementary particle because of the clumsiness of our equipment, the particle may still have well-defined properties; this position is untenable today.

Also, in his discussion of the 'quantum jungle', Stannard introduces the famous



double-slit diffraction experiment using the analogy of gazelles which, escaping from an attacking lion through two gaps in a hedge, quantum-diffract and end up in front of lionesses waiting at the positions where the interference maxima are to be expected. This analogy, attractive though it is, is inaccurate in that Stannard says the gazelles have no choice other than to act as they do because their dynamics is governed by quantum mechanics. This is incorrect, because the gazelles merely need to approach the 'double-slit' hedge at a slightly different angle, so that their interference maxima lie between the waiting lionesses, in order to escape.

Actually, from the point of view of modern physics, it is completely impossible to create a detailed visual description of what happens in a quantum experiment between the preparation of the system and the final observation. Therefore, even Gamow's drawings showing probability clouds or smeared-out tigers should be taken with a substantial grain of salt. Given their invalidity in principle, it might have been better to retain Gamow's original figures and thus, through their historic quaintness, imply that the conceptual picture is outdated, rather than replace them with new ones which still have their severe limitations.

That said, although presenting quantum mechanics accurately to a broad lay readership will certainly remain a challenge for some time, *The New World of Mr Tompkins* is a lovely book which I am sure everyone interested in modern physics, from the age of 11 upwards, will enjoy enormously.

Anton Zeilinger is at the Institut für Experimentalphysik, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria.

Hands-on taxonomy

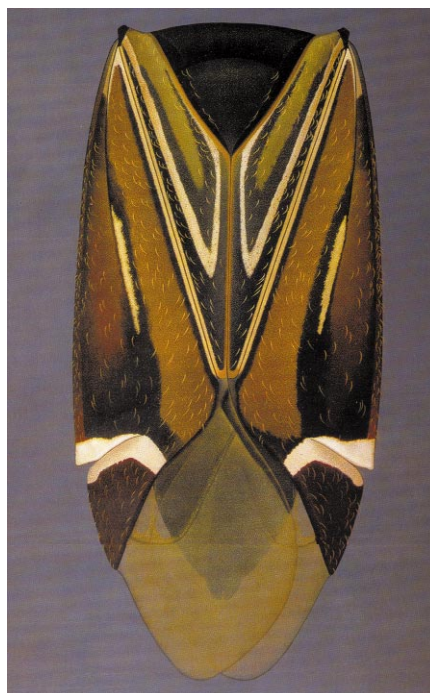
Describing Species: Practical Taxonomic Procedure for Biologists

by Judith E. Winston
Columbia University Press: 2000. 518 pp.
£11, \$35 (pbk)

Henry Disney

The discipline of taxonomy traditionally covers three areas: alpha, beta and gamma taxonomy. These areas respectively cover the recognition and description of species; the arrangement of species in classifications, which today aim to reflect phylogenetic affinity; and infraspecific categories such as subspecies, ecotypes and polymorphisms. There is currently considerable interest in beta taxonomy, particularly in using novel molecular data, and molecular techniques have also revived interest in problems concerned with gamma taxonomy. The bedrock discipline of alpha taxonomy, however, continues its relentless decline. Today, most specialists are retired professionals or amateurs, with the result that other biologists, such as ecologists, must frequently describe the novel species they encounter themselves.

With current estimates suggesting that around 90 per cent of the world's species remain undescribed, this situation will not soon change. Indeed, it reflects my own experience. Although primarily an ecologist/naturalist, even when working on well-studied insect groups as a medical entomologist, I was forced into describing new species. Now alpha taxonomy is my principal tool for investigating the natural histories of scuttle flies (Phoridae). Only about 25 per cent of my publications are straight ecology, whereas 19 per cent are ecology plus alpha taxonomy and 45 per cent are alpha taxonomy alone. Thus, my largest contribution has been in alpha taxonomy, even



A bug's world: Cornelia Hesse-Honegger's painting *Leaf Bug* (Miridae, *Cremnocephalus*).

though I have not been trained in this discipline and consider it only a means to an end.

A welcome arrival on this scene, Judith Winston's book is primarily concerned with alpha taxonomy. It is arranged in three parts — how to recognize that a specimen belongs to an undescribed species; how to prepare a description of the new species (including the choice of a name that conforms with the provisions of the relevant code of nomenclature); and how to deal with other problems, such as beta and gamma taxonomy, key construction and missing types.

The book is probably the most exhaustive treatment available of the practical aspects of describing new species or higher taxa. It enables the novice, such as an ecologist who finds that a species of interest is new to science, to prepare for publication an acceptable description of a hitherto undescribed

taxon. If I had had Judith Winston's book at the beginning of my career, I would not have fallen into so many of those holes — be they nomenclatural or scientific — that seem to litter the path of the would-be alpha taxonomist.

Henry Disney is in the Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

The Wonderland of primordial life

Life Without Genes

by Adrian Woolfson
HarperCollins: 2000. 409 pp, £17.99

John Godfrey

Complex features of organisms have, since Darwin, challenged anyone thinking deeply about evolutionary explanation. The vertebrate eye, or the interaction between the physiology and behaviour of a female mammal and her offspring during lactation, have obvious adaptive value when, but only when, fully developed. J. B. S. Haldane and others have argued that each stage of the origin of such features must have conferred selective advantage, with the finesse that the primordial stages may often have served a different function from the present one. These explanations are essentially darwinian, relying, as they do, on variation and natural selection. They apparently fail to account for the origin of the thing that varies: the integrated system of proteins, DNA, transcription and repair that characterize living cells.

Adrian Woolfson argues that earliest life forms had limited inheritance and that this was analogue in nature, rather than the precise digital kind that DNA now provides. This liquid life without clear boundaries and lacking genes would have been unstable and evanescent. He sets himself the task of showing that there are many possible paths from there to here, and also that the future, too, has many options. He makes a good case that these geneless entities may have been parasitized by digital proto-genes that took over their metabolism. The parasites were presumably polymers that encoded information in their combinatorial sequences. Ribosomes, too, may originally have been independent bacterial entities, now retaining their own common characteristics, including a few genes, even in current organisms of the most diverse kinds. Being shared by all, they must have been conservative for 3,000 million years or so, more, if cosmic dust imported life from space. Elsewhere, Woolfson discusses the genetic 'wild-type', but without mentioning that heterozygosity is the norm. The latter probably arose as a means of defence against parasitism and disease, as sex itself is likely to have done.

New in paperback

Mammals of the Neotropics, Volume 3: The Central Neotropics: Ecuador, Peru, Bolivia, Brazil

by John F. Eisenberg & Kent H. Redford
University of Chicago Press, \$28, £40

From White Dwarfs to Black Holes: The Legacy of S. Chandrasekhar

edited by G. Srinivasan
University of Chicago Press, \$18, £13

Wildlife of the Tibetan Steppe

by George B. Schaller
University of Chicago Press, \$20, £14

"Wildlife of the Tibetan Steppe will long remain a unique, important source of biological, but also sociologist, insights and challenges." Valerius Geist, *Nature* 395, 236 (1998)

Animal Revolution: Changing Attitudes Towards Speciesism

by Richard D. Ryder
Berg, £14.99, \$19.50

Ecology and Evolution of Darwin's Finches

by Peter R. Grant
Princeton University Press, £14.50, \$22.95

In a characteristically revealing episode, Woolfson outlines how molecules such as haemoglobin change their form reversibly to good adaptive effect in the lives of, for instance, penguins and crocodiles. The folding of haemoglobin makes its affinity for oxygen suit the changing physiological demands during diving. Elsewhere, he mentions that fungi and yeasts use a protein-based system of inheritance to a small extent. Perhaps this is a relict of a previously more important mode, ousted as determination of protein sequences by genes became the norm. If the protein-only theory of prion disease is indeed correct, then the strain variation in scrapie, for instance, must be based on fine changes of conformation that can pass from generation to generation of the host. This could be of great importance for the prospects of dealing with vCJD.

Woolfson has written a remarkable book, in a new literary form, to elucidate these problems, both for biologists and for a much wider audience. He carefully avoids obvious jargon. At the same time he coins a lot of his own terms as he leads readers towards fresh ways to comprehend the actual biological world by introducing them to possible alternative worlds. He does this by interspersing throughout the book episodes of science fiction, leading us like a modern Alice through dreamy wonderlands of potentiality. These imaginative excursions serve to depict many levels of complexity. We tour landscapes of whole animals, of RNA, DNA, information and more — though he does not include an epigenetic landscape, such as C. H. Waddington used as a metaphor for the action of genes during development. Each provides vast choice and illustrates how what we have had in evolution is but a tiny sample of what might have been. Few readers will fail to be helped out of the mental rut that tends to limit the imagination of most of us.

Suggesting experiments is far more difficult, and the author is less successful in this. Indeed, at one point he says: "We are obliged to sit back and wait for the appropriate experiments to be conducted". The book is, however, laced with unexpected curiosities. Explaining analogue memory, he tells how Puccini, auditioning singers for the part of Rodolfo in *La Bohème* and hearing Caruso for the first time, exclaimed, "Who sent you to me? God?"

Life Without Genes is excellent for certain readers, but for whom is not so clear. Some people may be annoyed by its stylistic idiosyncrasy. Others may think it would have been better if shorter; but many will be charmed by it and be grateful for the mental jerks it encourages us to take. Most will learn a lot, but may be irritated by the lack of an index, or even reference from the text to the extensive bibliography. HarperCollins, please note. ■

John Godfrey is at 41 Lawford Road, London NW5 2LG, UK.

Science in culture

A celebration of civilization

Seven Hills: Images and Signs of the 21st Century, an exhibition at the Martin-Gropius-Bau museum in Berlin, running until 29 October.

Alison Abbott

This is not the usual apocalyptic vision of science, nor is it the didactic, self-consciously 'entertaining' public-understanding-of-science display to which the German public is usually subject. This grandiose exhibition in Berlin places science and technology alongside the other aspects of culture that define our civilization, in a distinctly intellectual and sophisticated juxtaposition.

The 'Seven Hills' of the exhibition, Berlin's millennium project, are thematic installations created by the authors of our civilization — artists and architects, as well as scientists — and displayed in a nineteenth-century building in east Berlin.

The themes are big and somewhat abstract: nucleus, jungle, cosmos, civilization, knowledge, faith, dream. The overall aim is also big and somewhat abstract, portraying the essence of our social evolution. Each installation is designed by a different architect and different teams of scientists. Ken Adam, the Berlin-born film architect best known for creating the style of James Bond films and *Dr Strangelove*, has made in the building's central atrium a cathedral whose windows are represented by a five-metre-diameter particle detector suspended below the atrium's glass dome and above a pyramid of displayed objects, many shown from their inside view. The main focus is a globe whose burning interior erupts through magma canals to the surface. Below the globe, robot dogs and the Japanese P3 humanoid — the world's most advanced robot — stalk the busy floor, around DNA sequencers and other high-tech hardware. Other displayed artefacts, for example the skull of philosopher René Descartes and the brain of the nineteenth-century biologist Ernst Haeckel, serve as reminders that these technical advances are but the products of the human mind, individually ephemeral but feeding the collective memory that is our civilization.

The message conveyed by the Jungle installation is that nature is no longer natural, but bends to the will of our culture, even in matters of conservation. The immense statue of Athena, on loan from the neighbouring Pergamon museum, stonily monitors visitors to the Knowledge installation, with its displays of the artefacts that have been used to record knowledge, from parchments to computers, and those representing the political and religious institutions that control knowledge and its acquisition.

The architectural design of the Faith installation is a huge sphere splintered into ever-smaller fragments symbolizing the fragmentation of religions during the history of civilization. The Dream installation, which deals with subjectivity,



The glowing globe, part of the exhibition's 'Nucleus' installation.

the preserve of both artists and neuroscientists, is housed in a theatrical set of rooms created by Japanese stage designer Kazuko Watanabe.

The work of artists from every century cuts through each installation. A huge seventeenth-century painting by Johann Melchior Roos, *The Menagerie of Landgrave Carl von Hessen-Kassel*, newly restored for the exhibition, opens the Jungle installation, which also displays sixteenth-century watercolours by naturalist Giorgio Liberale, and the shocking works of the contemporary artists Jan Fabre, who makes sculptures from beetles, and Cornelia Hesse-Honegger, whose drawings record the anatomical deformations of bugs collected from around nuclear power plants.

Most of the installations feature hands-on tricks. In Jungle, one can interact with a sculpture of the Indian goddess Kali, viewing fantastic worlds literally through her eyes; in Space one can walk through a revolving tunnel that conveys a feeling of weightlessness.

But the exhibition is not play. Nor does it attempt to aestheticize science. Rather it integrates science into a complete aesthetic representation of civilization. The whole is also an optimistic statement about Berlin, which seeks to integrate its torn twentieth-century history into world civilization, the metaphoric Seven Hills, and into a positive future. This optimism is also represented by the participation of Ken Adam, a Jew forced out of Berlin during the rise of Nazism in the 1930s. Those Berliners rooted more prosaically in the present see not the optimism, but the DM28 million price tag presented to the reunified, but bankrupt, city. ■

Alison Abbott is the senior European correspondent of Nature.

Show them how it's *really* done

The public's appetite for science will not be whetted by a diet of dry facts.

Nancy J. Rothwell

Peter Medawar, Nobel laureate, was remarkably perceptive: "the ideas of the educated lay public on the nature of scientific enquiry and the intellectual character of those who carry it out are in a state of dignified, yet utter, confusion. Most of these misconceptions are harmless enough, but some are mischievous, and all help to estrange the sciences from the humanities and the so called 'pure' sciences from the applied," (*Encounter*, August 1965). Medawar had put his finger on a major issue — and one that is probably even more relevant today than it was 35 years ago.

Science has never had a higher public profile. The public understanding of science is hotly debated in most of the developed world, as governments, educationalists and scientists themselves realize the need to explain what scientists do and the benefits that accrue from their (often expensive) labours. But much of this misses Medawar's key points. Knowing scientific facts does not automatically translate into an understanding of science. Sadly, few outside the profession know what scientists actually do, or what drives them to do it. Indeed, most know little of what science is really about.

The communication of science to society is not new, nor is it an easy task. It is almost 200 years since Albemarle Street in London, home of the Royal Institution, was designated the first one-way thoroughfare — because of the crowds that flocked to see Humphrey

Davy's lectures. The remarkable popularity of these lectures and those of his successor, Michael Faraday, was probably because they used fascinating demonstrations, rather than simply telling people what to believe. Davy and Faraday were also great storytellers — and stories, of course, are the key to captivating an audience.

Some of the best accounts of science tell of people as much as facts, of personal circumstances, battles with dogma and establishment, luck, despair, breakthroughs, and even a few scandals. The popularity of *The Double Helix* owed as much to James Watson's juicy descriptions of personal conflicts as to the magnitude of the scientific discovery. John Gribbin brought the complexities of defining the Hubble constant to a wide audience in his book *The Birth of Time: How We Measured the Universe*, and the story of Faraday's humble beginnings as the son of a Yorkshire smith has fascinated and inspired many.

Such accounts not only explain the complexities of science, but also give an insight into what science and scientists are really about. Today, the teaching of science in schools, and its presentation to the public, focuses heavily on facts, leaving little room for enquiry or imagination. This does not reflect the reality of research, and ignores the most exciting part — the process of discovery itself. Research is like a detective story, in which knowing 'whodunnit', and how, is not enough — it's the process of deduction that really appeals.

We should not underestimate the public's

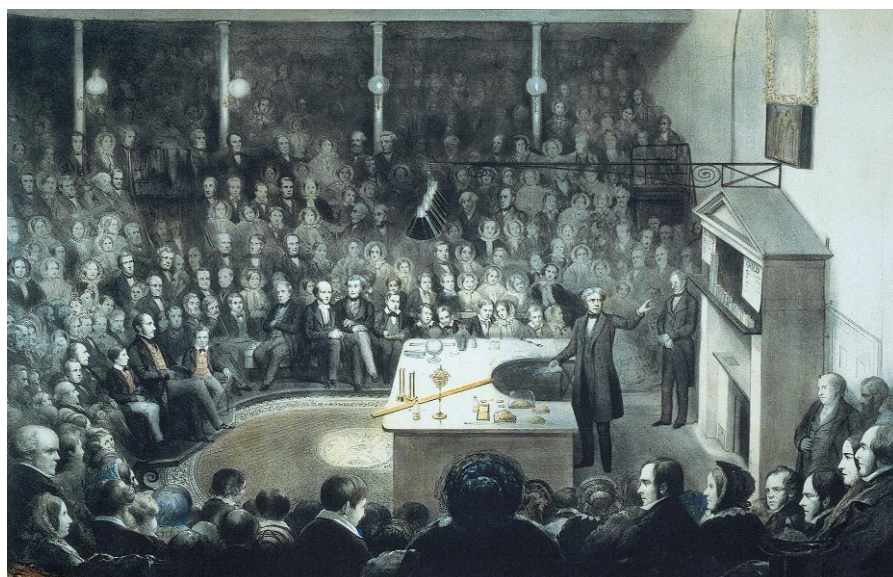
interest in science, nor its ability to comprehend complex issues. Fascination with the physical and natural world is not the exclusive domain of professional scientists, nor is it dependent on a higher degree or scientific training. But the playful yet influential portrayal by the popular media of the nutty professor or mad scientist does little to benefit science. Scientists themselves tend to promote logic, analysis and rigorous experimentation as the major (or even only) bases of research and discovery, when in reality they also rely on intuition, imagination and inventiveness, much of which does not follow the general view of common sense.

Medawar's most important point is one of the hardest to address — the growing divide between sciences and humanities, and between curiosity-driven and applied research. In presenting our work and discoveries to the public, we strive to justify their importance and cost by their benefits to humans, animals or the environment. And indeed, most people are supportive of research with obvious benefits, although this is like an investor telling his or her broker to buy shares — but only those that will increase in value.

To engage the public in what research is really about — the pursuit of knowledge — is a challenge, given science's growing complexity and rapid advance. Most advances have flowed not from a desire to improve society, but from natural human curiosity — although the benefits, sometimes not realized for many decades, have been considerable. The biologist J. B. S. Haldane believed that no one could have predicted that measuring the length of mercury columns could lead to an understanding of thunderstorms and fever, and the statistician Karl Pearson thought that Heinrich Hertz's discovery of electromagnetic forces had no useful application. Marconi proved him wrong.

Even Ernest Rutherford saw atomic physics as "useless" — at least initially. In the past decade, studies of the soil nematode *Caenorhabditis elegans* have revolutionized our understanding of cell death and survival, yet the worm biologists were probably not driven by the desire to understand or treat cancer and degenerative diseases, but rather by a fascinating biological problem. Discovery is the real pleasure of science. To give the last word to Medawar: "Pure science requires no justification outside itself, and its usefulness has no bearing on its valuation." ■

Nancy J. Rothwell is in the School of Biological Sciences, Stopford Building, University of Manchester, Manchester M13 9PT, UK.



Packing them in: Michael Faraday, with his storytelling and demonstrations, was a born showman.

New ice age, or just cold feet?

Temperatures boil over at a heated meeting on global cooling.

Norman Spinrad

Orbital Hilton, 14 March 2322. The highly touted First World Conference of Planetary Climatologists and Climatech Engineers on Global Cooling broke up in angry chaos on its opening day and fisticuffs were only averted by the quick action of hotel security staff, who pumped pax gas into the conference chamber before blows could be struck. It had been hoped that the neutral orbital venue would calm tempers, but unfortunately this did not prove to be the case.

The conference opened with a speech by Dr Vladimir Bunin, chairperson of the Committee of Concerned Climatologists, rehashing the charge that the past three centuries of effort by climatech engineers to counter the twentieth and twenty-first centuries' admittedly disastrous greenhouse warming had proven all too successful, and their meddling, if not speedily countered, was about to bring on a new ice age.

Bunin contended that the recent unplanned expansion of both polar ice caps, the increase in glaciation in the northern hemisphere, the return of winter snowfall as far south as Labrador and Moscow, and the icebergs sighted in the Bering Sea and off Tierra del Fuego were not — as the climatech engineers contend — isolated anomalies, easily corrected by adjustments to orbital mirrors and occluders and a diminished output by cloud-cover generators. Instead, he said, they were aspects of a dangerous global pattern leading to a new ice age that would devastate the agricultural breadbaskets of Siberia, Alaska, Canada and Lapland, upon which the Earth's 20 billion inhabitants depend.

He demanded the halving of the area of desert covered with mylar mirroring to lower the Earth's albedo and a redeployment of the orbital mirrors now sustaining the Gulf Stream to add more heat to the atmosphere. But the most heat was generated by his attack on Qwik-grow trees and the forestry programme.

"Far too many Qwik-grow trees have been planted, and now they are spreading like weeds, completely out of control," he declared. "It was one thing to reforest the Amazon basin, the American plains, the Indonesian archipelago and Europe while we were still burning fossil fuels, but now that our energy sources are all solar and thermonuclear, these vast and ever-expanding woodlands have created a severe atmospheric carbon dioxide deficit that has seriously

diminished the greenhouse effect. This can only be remedied by a return to the mass burning of coal or the setting of immense forest fires — preferably both."

This resulted not only in cries of outrage from the climatech engineer contingent, but the throwing of plastic coffee bulbs, styli, fruit, and other debris in his direction. Injury was prevented by the zero-g conditions, which rendered this fusillade ineffective, but the frustration only added to the ire of the climatech engineers, and it was a good 20 minutes before enough order could be restored for Hans Goodkin of Climatech Solutions to deliver a scathing rebuttal.



"Burn coal! Why not bring back the internal combustion engine, and to hell with the ozone layer! Or better yet, the steam locomotive, and never mind the acid rain! Burn the forests? Why settle for half-measures? Let's just nuke the trees! And while we're at it, why not drop thermonuclear charges into the craters of active volcanoes? The resulting mass eruptions would pump plenty of carbon dioxide back into the atmosphere and the red-hot lava flows would really heat things back up! Let's bring back Vesuvius and Mount St Helens and Krakatoa!"

When the booing and hissing subsided, he continued in an even more sarcastic vein: "Dissolving the coral reefs we've rebuilt would release plenty of carbon dioxide too! And we could melt the ice-caps from orbit! And replace the mylar mirroring with black ash! After all, if we listen to you people, we'll have plenty of soot to spare!" Unwilling to wait for the uproar to die down, Goodkin just turned up the gain on his microphone to override it.

"Let Central Australia and the Sahara savannah once more become howling deserts — no problem, their millions of inhabitants can join the rest of the planet's new boat-people when the sea levels rise again and the world's reclaimed coastlands and islands return from whence they came!"

Improvised missiles began to fly again — this time from the climatologists — or rather drift ineffectually in the general direction of Goodkin in the weightless conference chamber. "You're turning the Earth into a deep freeze to line your own pockets!" someone shouted.

"You climatological Luddite snake-oil salesmen want us to turn it back into a greenhouse oven so you can make your fortunes selling air-conditioners in Antarctica and ice-boxes to Eskimos!" Goodkin shouted back.

At which point, the Concerned Climatologists attempted to storm the podium, swimming clumsily towards it through the air like a school of slow-motion sharks, and hotel security released the pax gas.

"What do a climatech engineer, a climatologist, a planet and a nymphomaniac have in common?" a reporter asked, several drinks later, in the bar.

"I'm afraid you're about to tell us," groaned another.

"They're much easier to heat up than cool down."

Norman Spinrad is a novelist, screenwriter, essayist, songwriter and sometime performer. His latest novel, *Greenhouse Summer*, is published by Tor Books.

The realms of Archaean life

Euan Nisbet

There is evidence of a variety of early organisms from the Archaean — some 4,000 to 2,500 million years ago. Now life at deep-sea hot springs can provisionally be added to the list.

Beginnings; are apt to be shadowy" wrote Rachel Carson¹. So it is with the early record of life. In the past few years, plenty of clues have emerged, both from geological studies and inferences drawn from the molecular biology of modern organisms. Often the veil of shadows has lifted a little, only to show more darkness behind. Yet, bit by bit, speculation as to the nature and ecology of early life is being based on a firmer footing.

Rasmussen (page 676 of this issue²) is the latest to tweak the veil — he reports evidence of early microbial life in deep-sea volcanic rocks that are some 3,200 million years old. The evidence is thread-like filaments, twining and twisting in different directions, that may be fossils of microorganisms. The filaments occur in a volcanogenic massive sulphide. This is a type of metal deposit that is usually formed on the sea floor, at depths far below the light-penetration zone, and at temperatures near or above the limits tolerated by life.

Rasmussen interprets the filaments as the remains of subsurface prokaryotes, microbes such as bacteria that inhabited environments beneath the sea floor. The implication is that the organisms were chemotrophic hyperthermophiles — that is, non-photosynthetic organisms, which used inorganic matter in their environment as an energy source, and which lived at or near 100 °C. Prokaryotes spread readily: if Rasmussen's interpretation is correct, they would have been widespread in the deep-water volcanic systems of the time, living in and around hydrothermal systems (submarine hot springs).

The ecological map of the Archaean aeon, the time from around 4,000 million to 2,500 million years ago, is gradually becoming less vague. There is widespread evidence for photosynthetic life (both anoxygenic and oxygenic) before 3,000 million years ago, and possibly from as early as 3,500 to 3,700 million years ago³. There is now molecular fossil evidence for the existence of cyanobacteria 2,700 million years ago^{4,5}. Similarly, data on ancient carbon and sulphur isotopes^{6,7} support the deduction that methane-generating organisms — methanogens — recycled dead organic matter in sediment; and also that a diverse, hydrothermal-associated, microbial community of chemotrophs existed in waters of shallow and medium depths. There has

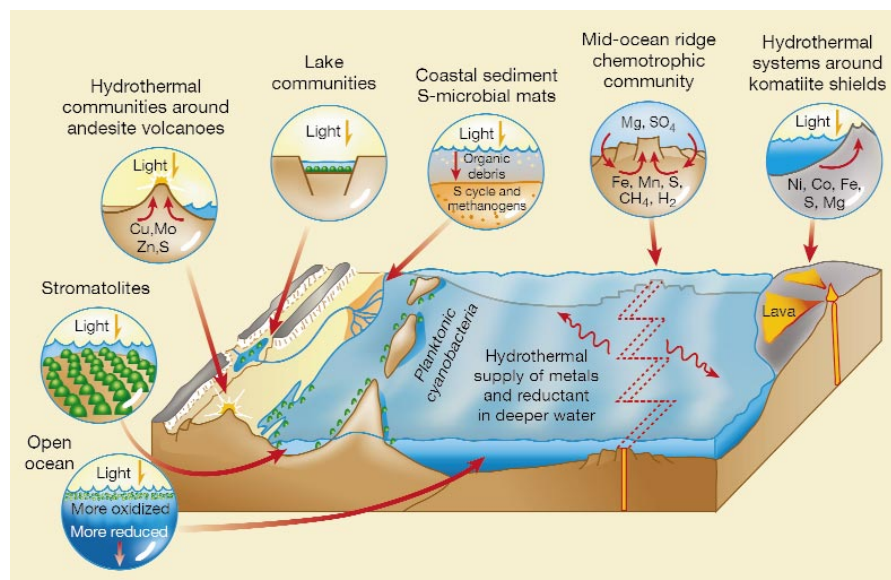


Figure 1 A map of Archaean ecology — places where early life may have flourished⁹. Among such settings are mid-ocean ridges (chemotrophic communities, including hyperthermophiles); hydrothermal systems around so-called komatiite shields and andesitic island-arc volcanoes (mixed chemotrophism and photosynthesis); shallow waters (oxygenic and anoxygenic photosynthesis); the open ocean (photosynthetic plankton); and muds in the sea floor (methanogens and exploitation of the sulphur cycle). New findings² also point to deep-water hydrothermal systems as habitats for life. In each habitat depicted here, the red arrows show fluid movement. The atmosphere at this time would have been rich in CO₂ and N₂, with minor amounts of NO_x, SO_x and CH₄.

been speculation that non-photosynthetic Archaean organisms lived in or around hydrothermal systems in deeper waters. But until the findings of Rasmussen² that view rested mainly on inference and on the surmise that certain deposits, interpreted as being biological in origin, formed below the base of wave action.

Physical models of the late Archaean Earth point to a variety of habitats for life (Fig. 1). In shallow coastal waters, there were microbial-mat communities, built by cyanobacteria, which carried out oxygenic photosynthesis. In muds and lower parts of the microbial mats, anoxygenic photosynthesizers and methanogens were probably present. Photosynthetic plankton were probably abundant in the ocean surface waters.

Rubisco is the enzyme responsible for carbon dioxide fixation in photosynthesis — isotopic evidence⁶ implies that rubisco-based bacteria directly handled about a fifth of the carbon flux emitted from the Earth's interior, as today, and by doing so managed the global carbon cycle. For life to be productive enough to manage carbon, it is likely that

modern-style sulphur and nitrogen cycles were also in operation.

Non-photosynthetic habitats, where sulphur chemistry could be exploited, were probably also widespread^{7,8}. Here, the communities could have depended on the contrast between relatively oxidized water and relatively reduced hydrothermal fluids, or they might have used reduced organic debris. In particular, hydrothermal microbial communities may have lived around mid-ocean ridges, volcanic ocean islands and island-arc volcanoes. In the mid and late Archaean, when oxidation power in the surface water was available from surface photosynthesis, such communities may have been prolific. Before the evolution of photosynthesis, the oxidation contrast would have been less. But it would nonetheless have been present, supplied as seawater sulphate.

The work of Rasmussen² and of others such as Roger Buick adds a new realm — predicted, but previously only surmised — to the map of Archaean ecology. Moreover, although Rasmussen's work does not show that deep-water hydrothermal life came

before photosynthetic life, it does lend circumstantial support to the argument that steps in the early history of life took place around hydrothermal systems. Analysis of ribosomal RNA in modern organisms allows evolutionary trees to be constructed, and inferences to be drawn from them. The 'standard' interpretation of this line of evidence⁸ is that chemotrophic hyperthermophile communities preceded the advent of photosynthesis. Perhaps, then, the earliest Archaean life was confined to the vicinity of hydrothermal systems.

Several implications follow from this possibility. Much of life's 'housekeeping' chemistry might have evolved in such settings, where metal sulphides were common and the danger of heat shock was ever present⁹. So, for example, proteins incorporating nickel may reflect their origin near high-temperature lavas (known as komatiites), which can contain nickel sulphide deposits; the use in proteins of copper, zinc and molybdenum (metals typical of hydrothermal deposits) could likewise record their heritage. Even the role of heat-shock proteins as chaperonins, which have a crucial part in assembling complex proteins, may date from the hazardous environment near hydrothermal systems.

Conceivably, sensitivity to temperature may also have developed around deep-water hydrothermal systems. This may have pre-adapted organisms to use light: spreading into shallower habitats, microbial mats may have exploited first anoxygenic then oxygenic photosynthesis in developing ecosystems of progressive complexity¹⁰.

Here we are deep into guesswork, however, and there are difficulties with the evidence. Disputes surround the interpretation of data provided by ribosomal RNA¹¹. And the history of the sulphur cycle is puzzling. Conventional isotope work suggests that there is little isotopic fractionation of sulphur in Archaean sediments, although such fractionation would be expected if complex sulphur cycles had evolved. Yet, from the molecular evidence it seems that the sulphur cycle, and by implication sulphur fractionation, is of great antiquity. However, recent high-resolution studies⁷ show a wide range of isotope ratios in well-preserved sediments 2,700 million years old.

The 'sulphur problem' may be related to the movement of fluid and sediment by multicellular organisms. Although there is geological evidence suggesting that ancestors to such organisms were present in the Archaean⁴, it may have been much later before they were physically able to control redox zones in microbial mats and produce regions of fractionated sulphur large enough to be sampled by conventional techniques.

Finally, the reasons for the rise in molecular oxygen, which seems to have occurred in the early Proterozoic some time before 2,000 million years ago, remain unclear: that rise

may have been linked to internal controls in the system incorporating rubisco, mitochondria and chloroplasts^{12,13}. And we can only guess about the history of predation.

The study of Archaean rocks is like a forensic investigation based on heavily smudged fingerprints. Rocks can be altered, whereas textures (such as reported by Rasmussen²) can be misinterpreted, and chemical and isotopic tracers^{4,5} can be of later origin. Yet, for all the uncertainties, the map of life's early evolution grows ever more detailed as new terrains are inked in. But, as Roger Buick has remarked, the map is still medieval in aspect, with large blank areas marked 'Terra incognita', or 'Here be cannibals'. And we still know little about the far edge of the world of life — where did it come from? ■

Euan Nisbet is in the Department of Geology, Royal

Holloway, University of London, Egham, Surrey TW20 0EX, UK.

1. Carson, R. *The Sea Around Us* (Penguin, London, 1956).
2. Rasmussen, B. *Nature* **405**, 676–679 (2000).
3. Rosing, M. T. *Science* **283**, 674–676 (1999).
4. Brocks, J. J. *et al. Science* **285**, 1033–1036 (1999).
5. Summons, R. E. *et al. Nature* **400**, 554–556 (1999).
6. Schidlowski, M. & Aharon, P. in *Early Organic Evolution: Implications for Mineral and Energy Resources* (ed. Schidlowski, M.) 133–146 (Springer, Berlin, 1992).
7. Grassineau, N. *et al. Terra Abstr.* **4**, 261 (1999).
8. Stetter, K. O. in *Evolution of Hydrothermal Ecosystems on Earth (and Mars)* (eds Bock, G. R. & Goode, J. A.) 1–18 (Wiley, Chichester, 1996).
9. Nisbet, E. G. & Fowler, C. M. R. *Geol. Soc. Lond. Spec. Publ.* **118**, 239–251 (1996).
10. Nisbet, E. G. & Fowler, C. M. R. *Proc. R. Soc. Lond. B* **266**, 2375–2382 (1999).
11. Doolittle, W. F. *Sci. Am.* 72–77 (Feb. 2000).
12. Joshi, H. M. & Tabita, F. R. *Proc. Natl Acad. Sci. USA* **93**, 14515–14520 (1996).
13. Tolbert, N. E. in *Regulation of Atmospheric CO₂ and O₂ by Photosynthetic Carbon Metabolism* (eds Tolbert, N. E. & Preiss, J.) 8–33 (Oxford Univ. Press, 1994).

Materials chemistry

Semiconductors meet biology

Chad A. Mirkin and T. Andrew Taton

Biomolecules are some of the most underused, yet powerful and versatile, building blocks available to the materials chemist. The molecular recognition properties of proteins, protein building blocks (peptides), and single-stranded DNA are unmatched by conventional synthetic analogues. Nature does not use biomolecules to recognize technologically important inorganic materials, but through advances in molecular biology there are now strategies for designing peptides to bind selectively to metal and metal-oxide surfaces. But many questions remain about the interactions between biomolecules and unnatural inorganic materials, and how their selectivity can be improved.

On page 665 of this issue, Belcher and co-workers¹ describe how they screen a combinatorial library, consisting of millions of peptides with unknown binding properties, to identify those peptides with selective affinities for various semiconductor surfaces. This approach has allowed them to select peptide sequences that can distinguish between different semiconductors, and even between specific crystalline faces of the same semiconductor.

The exquisite recognition properties of biomolecules are a consequence, in part, of their evolutionary refinement over millions of years. The use of such molecules in organizing non-biological inorganic objects (for example, tiny particles of insulators, semiconductors and metals) into functional materials is an important scientific frontier for the materials chemist in the twenty-first century. The ultimate goal is to identify a

type of highly specific binding property that can be used in the biologically driven assembly of molecular materials into functional bioinorganic structures.

At present, there are two approaches to exploiting biomolecules as the 'glue' for inorganic building blocks (Fig. 1). The first involves enhancing inorganic materials with biomolecules that have established and well-defined recognition properties, and then using these properties to generate and organize novel, hybrid materials (Fig. 1a). This strategy, which initially exploited the simple base-pairing interactions of DNA to assemble inorganic nanoparticles into periodic macroscopic structures², marked the birth of a new field focusing on designer materials whose physical and chemical properties are tailored by the choice of building blocks. The strategy has grown to include interactions between antibodies and antigens³, and between protein receptors and their ligands⁴, among the list of highly specific biomolecules.

This approach has led to the discovery of some fundamentally interesting and unique properties of these hybrid materials that make them particularly attractive for use in the fields of biodiagnostics and molecule-based electronics (or nanoelectronics). In the case of biodiagnostics, these properties include unusually sharp melting profiles, which translate into exceptionally high selectivities for the target material^{5,6}, optical properties that can be controlled by the choice of biomolecule⁷ or inorganic material⁸, and new catalytic properties that can be used in detection amplification methods.

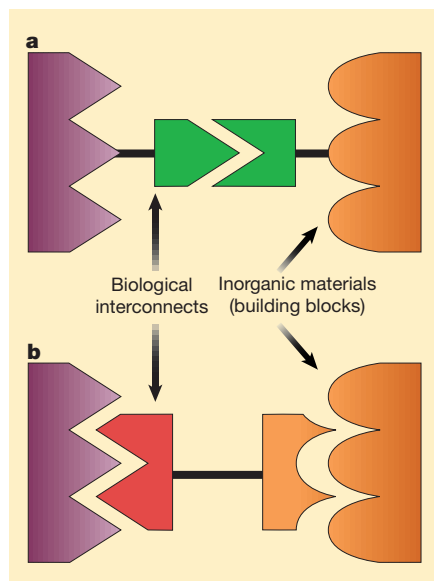


Figure 1 Two approaches to using biomolecules and their recognition properties for assembling non-biological inorganic materials into complex structures. **a**, Assembly of novel inorganic materials by 'programmed' interactions between biomolecules with known recognition properties. In this case, two different inorganic building blocks are modified with proteins, antibodies, oligonucleotides or ligands, which specifically recognize one another, allowing the formation of a hybrid material. **b**, Assembly of novel inorganic materials by tailored evolution of interactions between biomolecules and inorganic surfaces. In the study by Belcher and co-workers¹, the precise interaction of the biomolecular recognition element with the inorganic surface has been selected for, rather than explicitly designed. But this technique can still be used to link inorganic building blocks to form hybrid, designer materials.

In terms of molecule-based electronics, many view this approach as a route to biologically driven and programmed assembly of components that may be used by the semiconductor or optoelectronic industries. Such assembly processes could arrange nanoscopic materials in a massively parallel fashion and on a scale beyond conventional physical deposition techniques. Moreover, they provide excellent spatial control over the placement of the inorganic components with respect to one another^{9,10}.

The second approach uses the binding properties of biomolecules to recognize the inorganic materials directly (Fig. 1b). This is a more difficult technique because such recognition properties are not well known and must first be identified before they can be used in a rational way. Some organisms already use this principle for their own ends. For example, some proteins of the abalone shellfish provide a template for the crystallization of calcite into a hard shell¹¹, and antifreeze proteins prevent the growth of ice crystals in arctic fish¹². Several

groups have begun to explore the fundamental biospecific interactions involved in such processes to learn how to control them for engineering materials with new structures¹³.

Belcher and colleagues¹ have taken this principle a step further by developing a clever combinatorial approach to discover a family of biomolecules that recognize inorganic materials not ordinarily used by nature. Even more remarkable, the authors have used a trick of nature — natural selection of fragments from peptide libraries — to engineer these biomolecular recognition elements. In principle, this approach applies to all materials, so they may have discovered a way of directly interfacing biomolecules with any inorganic structure. Moreover, different parts of the same biomolecule could be designed to selectively recognize and organize multiple inorganic building blocks, creating structures with even greater spatial control. Although the work is at a preliminary stage and practical applications are some way off, this approach is likely to

become a powerful technique for the design of materials in years to come.

Chad A. Mirkin and T. Andrew Taton are in the Department of Chemistry and Center for Nanofabrication and Molecular Self-Assembly, Northwestern University, Evanston, Illinois 60208, USA.

e-mail: camirkin@chem.northwestern.edu

1. Whaley, S. R., English, D. S., Hu, E. L., Barbara, P. F. & Belcher, A. M. *Nature* **405**, 665–668 (2000).
2. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. *Nature* **382**, 607–609 (1996).
3. Shenton, W., Davis, S. A. & Mann, S. *Adv. Mater.* **11**, 449–452 (1999).
4. Niemeyer, C. M., Bürger, W. & Peplies, J. *Angew. Chem. Int. Edn Engl.* **16**, 2265–2268 (1998).
5. Elghanian, R. *et al.* *Science* **277**, 1078–1081 (1997).
6. Storhoff, J. J., Elghanian, R., Mucic, R. C., Mirkin, C. A. & Letsinger, R. L. *J. Am. Chem. Soc.* **120**, 1959–1964 (1998).
7. Storhoff, J. J. *et al.* *J. Am. Chem. Soc.* **122**, 4640–4650 (2000).
8. Mitchell, G. P., Mirkin, C. A. & Letsinger, R. L. *J. Am. Chem. Soc.* **121**, 8122–8123 (1999).
9. Mucic, R. C., Storhoff, J. J., Mirkin, C. A. & Letsinger, R. L. *J. Am. Chem. Soc.* **120**, 12674–12675 (1998).
10. Alivisatos, A. P. *et al.* *Nature* **382**, 609–611 (1996).
11. Belcher, A. M. *et al.* *Nature* **381**, 56–58 (1996).
12. Jia, Z., DeLuca, C. I., Chao, H. & Davies, P. L. *Nature* **384**, 285–288 (1996).
13. Perl-Treves, D., Kessler, N., Izhaky, D. & Addadi, L. *Chem. Biol.* **3**, 567–577 (1996).

Alzheimer's disease

Closing in on γ -secretase

Bart De Strooper

For decades a short amyloid peptide known as A β has been invoked to explain the neurodegeneration characteristic of Alzheimer's disease. This peptide is generated from the amyloid precursor protein by protein-cleaving (proteolytic) enzymes called β -secretase and γ -secretase¹ (Fig. 1, overleaf). Last autumn the β -secretase was identified, providing one molecular target for those seeking drugs to treat Alzheimer's disease^{2–5}. In their paper on page 689 of this issue⁶, Li *et al.* pinpoint the identity of γ -secretase. It turns out that presenilin, a suspect of old, is the culprit. Independent confirmation of this story comes in a paper by Esler *et al.*⁷ in the July issue of *Nature Cell Biology*.

Retrospectively, one wonders why it has taken so long to come to this conclusion. It is known that missense mutations in the genes encoding presenilin 1 and 2, which result in severe hereditary Alzheimer's disease, affect γ -secretase activity⁸. They cause a shift of two amino acids in γ -secretase's preferred cleavage site in the amyloid precursor protein (APP) sequence. This generates an amyloid peptide that is particularly prone to the precipitation and formation of amyloid plaques in the brain. A similar shift towards a longer A β peptide is observed with APP containing Alzheimer's disease mutations⁸. Apparently, this longer A β peptide is particularly toxic to the brain.

Until recently, presenilin was not considered to be a candidate protease (proteolytic enzyme), because of its highly unusual structure. Presenilin 1 and presenilin 2 are strongly hydrophobic and contain several membrane-spanning domains (Fig. 1). They have not the slightest similarity to any known protease⁹. But when presenilin 1 genes were inactivated in mouse neurons, γ -secretase activity and presenilin expression were both abolished¹⁰. And, in embryonic stem cells lacking both presenilin 1 and 2, cleavage of APP (and another protein, Notch) becomes undetectable^{11,12}. Moreover, Wolfe *et al.*¹³ found that the substitution of either of two aspartyl residues in transmembrane domains 6 and 7 of presenilin 1 was also sufficient to inactivate γ -secretase.

The bold conclusion that these aspartates constitute the active site of a completely new aspartyl protease¹³ now gains considerable experimental support from the work of Li *et al.*⁶ and Esler *et al.*⁷, who show that transition-state-specific γ -secretase inhibitors specifically bind to the presenilins. The inhibitors were modified with reactive groups that allowed them to crosslink covalently to their target protease. These targets turned out to be presenilin 1 and presenilin 2. The binding could be competed with unmodified γ -secretase inhibitors, demonstrating the specificity of the

findings. Li *et al.*⁶ further showed that presenilin 1 is first synthesized as a zymogen (an inactive precursor protease), which does not bind the inhibitor, and becomes active only when it is cleaved into amino- and carboxy-terminal fragments¹⁴. Both fragments appear to contribute to the active site of presenilin because, depending on the orientation of the crosslinking moiety, the inhibitor binds to either of them⁶. This finding is consistent with the model proposed by Wolfe *et al.*¹³, because each fragment contributes one aspartate to the putative active site of γ -secretase (Fig. 1). So we now have compelling evidence that presenilins are examples of a new class of aspartyl proteases.

Is research into presenilin and γ -secretase now at an end? Certainly not, and much more work will be needed before we really understand how hydrolysis of peptide bonds in the hydrophobic cell membrane can occur. The possibility that the transmembrane domains of presenilin provide a hydrophilic pocket, allowing water to enter

the catalytic site, needs further scrutiny. The identity and role of other proteins, thought by many investigators to be associated with the presenilins in a multiprotein complex, also needs clarification. The putative catalytic site of presenilin is especially intriguing because it is completely divergent from the consensus site of classical aspartyl proteases.

The development of clinically useful γ -secretase inhibitors is likely to be long and tedious. One hurdle will be the delivery of such drugs to neurons in the brain. For instance, the inhibitors used by Li *et al.*⁶ and Esler *et al.*⁷ do not cross cell membranes efficiently, and will probably be of little use *in vivo*. The good news is that pharmaceutical companies now have another defined target for their drug-discovery machineries.

The insights of the current studies go beyond research into Alzheimer's disease and are highly relevant to cell biology. The presenilins are only the second class of eukaryotic proteases found to be involved in regulated intramembrane proteolysis — a process that causes the proteolytic release of the cytoplasmic domain of integral membrane proteins. These proteins then migrate to the nucleus to activate the transcription of genes that are pivotal in cellular

differentiation, the unfolded-protein stress response and cholesterol biosynthesis¹⁵. Presenilin is apparently involved not only in APP processing, but also in regulated intramembrane proteolysis of the Notch receptor, which is a central player in development, and probably of the Ire1 protein, which controls the unfolded-protein stress response^{9,15}. All in all, there remains much to do — in particular, the hunt is now on for other substrates of presenilin, alias γ -secretase, and for proteins that regulate its activity. ■

Bart De Strooper is at the Center for Human Genetics, Flanders Interuniversity Institute for Biotechnology, K. U. Leuven, B-3000, Belgium.
e-mail: bart.destrooper@med.kuleuven.ac.be

1. Haass, C. & Selkoe, D. J. *Cell* **75**, 1039–1042 (1993).
2. Hussain, I. *et al.* *Mol. Cell. Neurosci.* **14**, 419–427 (1999).
3. Sinha, S. *et al.* *Nature* **402**, 537–540 (1999).
4. Vassar, R. *et al.* *Science* **286**, 735–741 (1999).
5. Yan, R. *et al.* *Nature* **402**, 533–537 (1999).
6. Li, Y.-M. *et al.* *Nature* **405**, 689–694 (2000).
7. Esler, W. P. *et al.* *Nature Cell Biol.* (in the press).
8. Scheuner, D. *et al.* *Nature Med.* **2**, 864–870 (1996).
9. Haass, C. & De Strooper, B. *Science* **286**, 916–919 (1999).
10. De Strooper, B. *et al.* *Nature* **391**, 387–390 (1998).
11. Zhang, Z. *et al.* *Nature Cell Biol.* (in the press).
12. Herreman, A. *et al.* *Nature Cell Biol.* (in the press).
13. Wolfe, M. S. *et al.* *Nature* **398**, 513–517 (1999).
14. Thinakaran, G. *et al.* *Neuron* **17**, 181–190 (1996).
15. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. *Cell* **100**, 391–398 (2000).

Semiconductor physics

Half-matter, half-light amplifier

Yoshihisa Yamamoto

A small semiconductor chip generating coherent light waves is commonplace in modern life, from global telecommunication networks to personal compact disk players. This semiconductor light-emitting device is based on the principle of 'light amplification by stimulated emission of radiation' (laser). Quantum mechanically, matter can also be regarded as waves. There are two fundamental types of particle that can make matter waves: bosons and fermions. As a consequence of the Pauli exclusion principle, two fermions cannot occupy the same state, whereas bosons tend to occupy the same state. A laser is based on the fact that particles of light — photons — are also bosons. According to quantum mechanics, a composite particle consisting of an even number of fermions behaves as a bosonic matter wave. So we can imagine building a laser or amplifier for such a matter wave. Writing in *Physical Review Letters*, Savvidis *et al.*¹ describe how they amplify matter waves in semiconductors.

The composite particle amplified by Savvidis *et al.* is known as an exciton-polariton. An exciton is a bound state of a negatively charged electron and a positively charged hole, which is analogous to a hydrogen atom.

Semiconductors can create them by absorbing photons. Eventually the electron and hole in the exciton will recombine to produce a photon. But if the photons emitted by the excitons in the semiconductor are confined in a reflective microcavity, the excitons and photons are mixed together to create a quasiparticle called a polariton (Fig. 1, overleaf). This new particle is part light and part matter, and consequently the polariton mass is reduced by four orders of magnitude compared with the exciton mass.

Unlike a photon, a massive particle cannot be created from a vacuum. The total number of massive particles must be conserved. In order to amplify a matter wave, a reservoir of identical particles is needed. Moreover, a scattering mechanism to transfer a reservoir particle into the ground state is essential. If the particle is a pure boson without charge, for example a photon in vacuum, there is no such scattering mechanism. But composite bosons, such as excitons, scatter with each other owing to the fermionic nature of their constituents — electrons and holes. This non-bosonic feature of a composite particle is the gain mechanism for a matter-wave amplifier.

In the experiment of Savvidis *et al.*¹,

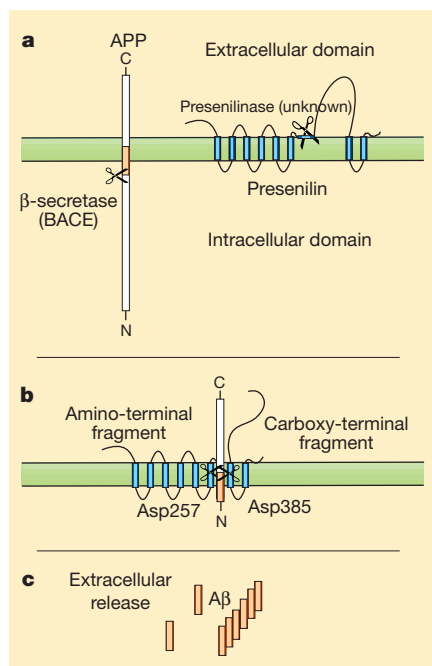


Figure 1 Generation of the amyloid peptide A β . a, Amyloid precursor protein (APP) is first cleaved by β -secretase, which releases its extracellular domain. Presenilin is processed shortly after its biosynthesis by an unknown protease, called presenilinase, to yield amino-terminal and carboxy-terminal fragments. b, These fragments remain non-covalently associated and form the active γ -secretase. Each fragment contributes one aspartyl residue (Asp 257 or Asp 385) to the active site. The β -secretase-cleaved APP fragment associates with and becomes cleaved by presenilin in an unidentified subcellular compartment. c, This results in the release of A β (orange) in the extracellular 'milieu'. A β then aggregates into the amyloid plaques observed in the brain of patients suffering from Alzheimer's disease.



100 YEARS AGO

The widespread invasion and persistent devastations of locusts in so many parts of Africa give interest to all trials and experiments, as well as the ordinary remedies, employed for the alleviation of this ruinous plague of the farmer. The following notes from Mr. W. C. Robbins, Stock Inspector of the Lower Tugela and Mapumulo Districts, are published in the Cape official *Agricultural Journal*:—"For the past three days I have been over the ground where my men have been infecting locusts with Government fungus, and the result was that I found dead locusts everywhere. I send you a sample; you will notice they are full of worms, and we know from experience that when locusts are found in this state whole swarms die off. Some you will see, are half eaten; these were eaten by their fellows. I have seen many clusters of locusts eating dead ones." The feeding upon bodies of dead locusts suggests that diseased locusts may be utilised as a substitute for locust fungus. From *Nature* 7 June 1900.

50 YEARS AGO

The gestural or 'ta-ta' theory of the origin of language has a long pedigree, beginning with Plato and achieving its most notable modern exponent in Sir Richard Paget. That Prof. Jóhannesson had been attracted to this theory was evident from his earlier work, "Um Frumtungu Indógermana og Frumheimkynni"; and the dedication of this new volume to Sir Richard sets the author decisively among those who believe themselves capable of demonstrating that language originated in the imitation by the vocal organs of physical shapes and movement. This theory represents one branch of the more general doctrine which sees a natural connexion between things and their names: and it is not to be denied that certain sound-complexes seem, in fact, to be particularly appropriate to the representation of certain shapes — a field of study in which Koehler and the *Gestalt* psychologists have collected some interesting data. But even supposing that these sound-complexes could be shown to involve vocal movements related to the shapes concerned, this would teach us nothing about the origin of language, which must be far remote from the attested or 'reconstructed' forms on which the theory is based — though Prof. Jóhannesson gives one the impression that he believes his hypothetical Indo-European forms at least to approximate to a primitive tongue. From *Nature* 10 June 1950.

two polaritons with the same quantum wavenumber are simultaneously created by a coherent laser pulse (the 'pump' pulse). These two polaritons are scattered to different energy states — the ground state and a higher-energy state — in a way that conserves energy and momentum between the initial and final states. When the ground-state polariton is stimulated by another coherent laser pulse, an enormous gain of up to 100 is observed.

This might seem like a matter-wave analogue of a laser amplifier. However, it does not correspond to a laser amplifier in an exact sense, but rather to a 'parametric' amplifier. In the radiowave, microwave and optical-wave spectrum, there are always two types of amplifier: a negative-conductance amplifier and a nonlinear-susceptance amplifier. A negative-conductance amplifier is an incoherent system, in which only the energy of the pump pulse determines the system dynamics. A laser amplifier is a classic example of a negative-conductance amplifier. A nonlinear-susceptance amplifier is a coherent system in which both the amplitude and the phase of the pump pulse determine the system dynamics. A parametric amplifier is a classic example of a nonlinear-susceptance amplifier. In the experiment of Savvidis *et al.*¹, the ground-state polariton is amplified by a phase-coherent parametric process.

In a different study, Huang *et al.*² create a direct matter-wave analogue of a laser amplifier using essentially the same structure as Savvidis *et al.* In this experiment, two reservoir polaritons with opposite quantum wavenumbers are scattered to the ground and excited states with zero wavenumber. Amplification is observed even after the reservoir polaritons lose the phase information of the pump laser and reach thermal equilibrium. This means that matter-wave amplification in this system could be achieved using electrically injected incoherent excitons.

Matter-wave amplifiers using atoms^{3,4} rather than excitons operate on the same principle, but their working conditions are quite different. The effective temperature of an exciton reservoir is about 10 K, whereas that of an atomic reservoir (Bose-Einstein condensate) is about 10 nK. The gain and loss rates of an exciton matter-wave amplifier are approximately 10^{12} s^{-1} , whereas those of an atomic matter-wave amplifier are approximately 10^5 s^{-1} . The ratio of the average spacing between two excitons compared to their radius is about 30, whereas that of the average spacing between two atoms to the atomic radius is about 10^3 . Exciton matter-wave amplifiers operate at a much higher temperature, gain and particle density than their atomic counterparts.

The exciton laser and amplifier can operate with a much smaller pump power than a normal photon laser. A photon laser and amplifier requires more populations in an

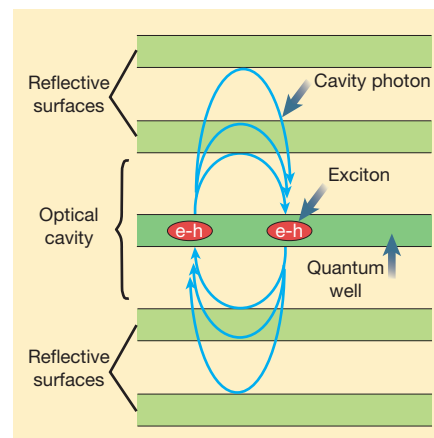


Figure 1 A semiconductor matter-wave amplifier. Here, a semiconductor quantum well is sandwiched between two reflective microcavities, which are separated by one-half or one wavelength. An exciton, which is a bound state of an electron-hole (e-h) pair and the solid-state analogue of a hydrogen atom, is confined in the quantum well by the potential barrier between the quantum well and the surrounding materials. A light field is trapped by multiple reflection from the reflective interfaces. When a quantum-well exciton and a cavity photon couple strongly with each other, they form a new quasiparticle called a polariton. Savvidis *et al.*¹ show that the polaritons can be excited resonantly by a laser producing amplification of the signal.

excited state than in the ground state (so-called population inversion). But the exciton laser and amplifier operate without inversion⁵. A wide bandgap semiconductor, such as gallium nitride or zinc oxide, is more suitable for constructing such an exciton laser and amplifier because the excitons in such materials are stable at higher temperatures and densities. Such semiconductor lasers are attractive because they are an efficient source of blue light.

If the gain of a matter-wave amplifier exceeds the loss rate, the system is expected to form a matter-wave laser (boson)⁵. But the spontaneous build-up of a coherent matter wave without continuous pumping by an input matter wave has yet to be seen. Photoluminescence studies of microcavities in semiconductor quantum wells^{6,7} suggest that this goal is within the reach of current technology.

Yoshihisa Yamamoto is at the Quantum Entanglement Project, Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305-4085, USA, and NTT Basic Research Laboratories, Atsugishi, Kanagawa, Japan. e-mail: yamamoto@loki.stanford.edu

1. Savvidis, P. *et al.* *Phys. Rev. Lett.* **84**, 1547–1550 (2000).
2. Huang, R. *et al.* *Phys. Rev. B* **61**, R7854–R7857 (2000).
3. Inouye, S. *et al.* *Nature* **402**, 641–644 (1999).
4. Helmerson, K. *Nature* **402**, 587–588 (1999).
5. Imamoglu, A. *et al.* *Phys. Rev. A* **53**, 4250–4253 (1996).
6. Dang, L. S. *et al.* *Phys. Rev. Lett.* **81**, 3920–3923 (1998).
7. Senellart, P. & Bloch, J. *Phys. Rev. Lett.* **82**, 1233–1236 (1999).

Laser-matter interactions

It takes two electrons to tango

Keith Burnett

Understanding the behaviour of electrons is fundamental to the science of matter and its properties. It usually relies on individual electrons adopting their own paths or 'orbitals', and explains the periodic table of the elements. But electrons are not always so aloof, and when they dance together they give rise to many intriguing physical phenomena, such as superconductivity. On page 658 of this issue, Weber *et al.*¹ provide a clearer picture of a two-electron dance in a different domain — when matter is exposed to an intense laser field. In so doing they have cleared up a controversy about how electron coupling occurs in such systems.

The interaction of matter with intense laser fields crops up in many areas, from laser surgery to ignition of nuclear fusion. Single-electron processes in atoms are fairly well understood, in both weak and strong fields. Systems of many electrons (which covers most of physics and chemistry) are much more complex, and there are fundamental questions to be answered. The study of matter in intense laser fields has been driven by developments in laser technology that make it possible to expose atoms to electric fields that rise to extraordinarily high values in a few femtoseconds (10^{-15} seconds). The peak field is as strong as that produced by the nucleus, and the electrons are ripped off the nucleus in short order. The way in which this apparently simple process — called ionization — actually happens has been challenging scientists in the fields of atomic, molecular and optical physics for some time.

Workers in these areas are now looking at multi-electron effects in atomic ions and clusters. One of the most intriguing of these is the 'double ionization' of helium in an intense laser field — that is, when both helium electrons are emitted at the same time. For certain values of the laser field the independent-electron picture fails spectacularly²: pairs of electrons are produced much more copiously than expected on the basis of the behaviour of independent electrons.

Theories put forward to explain the data give radically different pictures of what is going on. One of the most popular models proposes a mechanism known as 'rescattering', in which one electron is first accelerated away and then driven back towards the atom by the laser field, where it can collide with the remaining electron^{3–5}. Electrons emitted in this way would be expected to have highly correlated motion. Another model suggests that the second electron is shocked out of its orbit by the disappearance of the first electron⁶. In this case the motion of one is

not expected to depend on the motion of the other.

This controversy arose because the theory of strongly correlated electrons is such a tricky business. In terms of static properties of atoms, powerful techniques have given us a good picture of the correlated motion of electrons. In the case of intense laser fields, the problem is much harder. The system is evolving in time and the laser field destroys any simplification due to atomic symmetry. This means that the dynamics of even a single electron in an intense laser field is a serious computational problem. The availability of high-speed computers has made this an easier, but still not routine, task. More recently, the advent of massively parallel machines has made it possible to tackle the more difficult problem involving two electrons⁷. In that study, the emitted electrons appear on the same side of the atom, with correlated motion.

So the scene was set for an experiment to decide between the theories. The experiment

of Weber *et al.*¹ has confirmed that electron correlation is responsible for the anomalous production of electron pairs in an intense laser field. To see this, the authors have developed a technique that maps the distribution of electrons ripped out of an atom by an ultra-strong laser pulse. This enables the authors to determine the distribution of both single- and double-ionization events. The results clearly show a strong correlation between the momentum of the electrons. They also show that correlations dominate at particular laser intensities, which supports the rescattering mechanism of double ionization. This is a reminder that we have to get inside the complexities of the two-electron dance if we want to understand matter in intense fields.

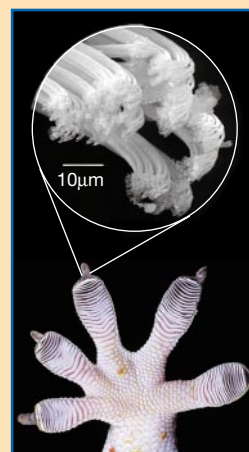
This work has given us a clearer picture of how multi-electron systems behave in strong laser fields. It should also give a great boost to their study in other systems, such as clusters of atoms or atomic surfaces. We now want to find out how electrons are correlated during the intense-field ionization and heating of atomic clusters. In particular, clusters can be rapidly heated by a laser field to produce small nuclear fusion bursts. Future experiments will challenge our theoretical understanding of interacting electrons in large assemblies of atoms (currently handled through density functional theory). The

Biomechanics

Gripping feat

Humid nights in cheap hotels in the tropics are not complete without a gecko scuttling up the walls or across the ceiling. The adhesive properties of the feet of these fleet lizards are proverbial, yet decades of investigation have still not revealed exactly how they do it. In their report elsewhere in this issue (*Nature* **405**, 681–685; 2000), Autumn *et al.* come the closest yet. Their force measurements on adhesive setae (foot-hairs) from the Tokay gecko, *Gekko gekko*, are consistent with the rapid formation and breaking of intermolecular bonds — van der Waals forces.

Suction and friction — two other hypotheses — have been ruled out, as gecko feet can work in a vacuum, and the lizards are quite content on surfaces as smooth as polished glass. Electrostatic attraction is also unlikely, as the feet can still adhere in ionized air. And there aren't any gland cells that might



produce some kind of glue. The role of adsorbed water, however, has yet to be studied.

Nevertheless, the observation that geckos get stickier with increasing surface energy of the substrate suggests that they are tapping directly into the molecular structure of the surfaces they walk on.

Viewed through a microscope, the foot of a gecko is densely packed with fine

setae, shown here in the inset. There are around 5,000 per square millimetre — around half a million on each foot. The end of each seta is further subdivided into between 400 and 1,000 structures called spatulae. The setae tend to point towards the heel: as the gecko takes a step, driving the sole into the substrate and pushing it backwards, the setae become maximally engaged. If all the setae were simultaneously stuck to the surface, the feet of a gecko could produce an adhesive force equivalent to ten atmospheres. The gecko releases each foot by 'peeling' off the setae at a critical angle, rather like peeling adhesive tape. Engineering a structure as exquisite as the foot of the gecko is probably beyond human technology, say Autumn *et al.*, but the principles on which it operates could inspire the design of new kinds of dry adhesive.

Henry Gee

MARK MOFFETT

present studies of electron correlation provide a benchmark for the development of methods that are used in many areas of physics and chemistry.

Keith Burnett is at the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

Structural biology

Pumping ions

David H. MacLennan and N. Michael Green

Calcium activates a variety of cellular responses when it enters the cytoplasm of a cell by means of transmembrane channels. But, to be effective as a signal, its concentration must be returned to sub-micromolar levels by ATP-driven pumps. The Ca^{2+} pump is a membrane-bound, Ca^{2+} -activated ATPase^{1,2}, similar to the Na^+/K^+ pump that controls ion balance and membrane potential in all animal cells³. These pumps belong to a superfamily of ATPases known as 'P-type', because they depend on the autophosphorylation — using ATP — of a conserved aspartic acid residue.

On page 647 of this issue, Toyoshima *et al.*⁴ describe the first high-resolution structure of any member of this family — the sarco(endo)plasmic reticulum Ca^{2+} -ATPase (SERCA1a). This important structure illuminates past investigations and reveals a model for the structural changes that underlie the activity of this pump.

In the forty years since its discovery, ingenious experiments have revealed much about how the Ca^{2+} -ATPase works. Kinetic studies⁵ showed that this pump transports Ca^{2+} by a reversible cycle (Fig. 1), but this vectorial transport can be understood only in a structural context. After the initial topology of the Ca^{2+} pump was deduced^{6,7} and refined⁸, electron microscopy⁹ provided a map of the molecule at a resolution of 8 Å. The structure of a homologous enzyme, haloacid dehalogenase, was recently modelled into the map¹⁰ to give a picture of the cytosolic phosphorylation domain that forms the catalytic heart of the Ca^{2+} pump. The two Ca^{2+} -binding sites were proposed to lie side by side near the centre of four transmembrane helices, and to be formed by the appropriate juxtaposition of acidic and oxygen-containing amino acids in a region of the protein that is sensitive to mutation^{11–13}. So, small changes in these helices could lead to disruption of the Ca^{2+} -binding sites and to alterations in the movement of ions to the cytosol or to luminal spaces. These studies made clear that the Ca^{2+} -binding sites in the membrane domain and the phosphorylated aspartic acid residue in the cytosolic domain

e-mail: k.burnett1@physics.oxford.ac.uk

1. Weber, Th. *et al.* *Nature* **405**, 658–661 (2000).
2. Walker, B. *et al.* *Phys. Rev. Lett.* **73**, 1227–1231 (1994).
3. Corkum, P. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
4. Becker, A. & Faisal, F. H. M. *J. Phys. B* **29**, L197–L202 (1996).
5. Watson, J. B. *et al.* *Phys. Rev. Lett.* **78**, 1884–1887 (1997).
6. Sheehy, B. *et al.* *Phys. Rev. A* **58**, 3942–3952 (1998).
7. Taylor, K. T. *et al.* *Laser Phys.* **9**, 98–104 (1999).

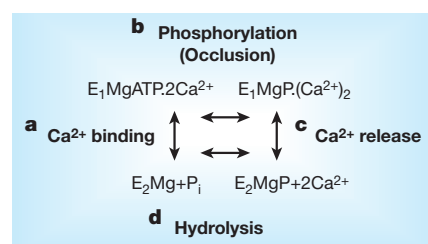


Figure 1 Conformational states of the Ca^{2+} -ATPase. During the cycle of ATP-dependent Ca^{2+} transport, at least four interconvertible phosphorylated and unphosphorylated conformations have been defined. a, A rise in Ca^{2+} on the cytoplasmic side of the membrane saturates the two high-affinity Ca^{2+} -binding sites to form $\text{E}_1\text{MgATP} \cdot 2\text{Ca}^{2+}$. b, This activates formation of a phosphoenzyme with MgATP, occluding Ca^{2+} within the protein to form the high-energy intermediate $\text{E}_1\text{MgP} \cdot (\text{Ca}^{2+})_2$. c, In a rate-limiting step to generate E_2MgP , the phosphoenzyme loses both its ability to re-phosphorylate ADP and its high affinity for Ca^{2+} , and opens its gate to luminal spaces. d, Water enters the catalytic site and hydrolyses the phosphorylated aspartic acid residue to regenerate the ATPase (E_2).

are separated by 40–50 Å, raising the question of how these long-distance interactions are mediated⁸.

The Ca^{2+} -bound structure described by Toyoshima *et al.*⁴ offers many more insights into the mechanism of action of Ca^{2+} -ATPase (Fig. 2, overleaf). The authors identify new ligands, particularly mutation-sensitive backbone carbonyl groups on transmembrane helix M4, and a single water molecule, in the Ca^{2+} -binding site, together with possible entry and exit pathways for Ca^{2+} that are lined by oxygen atoms. A striking feature of this region is the disruption of the M4 and M6 helices to form a Ca^{2+} -binding cavity. This was anticipated from earlier NMR studies of helix M6 (ref. 14), and means that structural changes in the carboxy termini of M4 and M6 may be the key to creating or closing off access to the Ca^{2+} -binding cavity.

Of the three cytosolic domains, the phosphorylation (P) and nucleotide-binding (N)

domains form the catalytic site, whereas the actuator (A) domain (formerly called the transducer or beta domain) appears to be involved in the transmission of major conformational changes. Toyoshima *et al.*⁴ confirm the structural similarity of the P domain to the catalytic domain of haloacid dehalogenase¹⁰. Note particularly that the positions of important catalytic residues (Fig. 2) are conserved.

The N domain, which is inserted, as predicted, after the first strand of the P domain (Fig. 2), forms an open cap over the catalytic site. In the Ca^{2+} -bound conformation⁴, the P and N domains appear as an open jaw, so the binding site for the ATP homologue TNP-AMP in the N domain is more than 25 Å from the critical aspartic acid, Asp 351, in the P domain. This implies that Ca^{2+} binding alone is insufficient to produce a phosphorylatable conformation of this protein. But the closure of the N and P domains must occur to bring ATP close to Asp 351, and possibly to exclude water and stabilize the phosphorylated enzyme. A cautious interpretation is required, however, as TNP-AMP differs from closer analogues of ATP, which cause conformational changes and disrupt crystals of the Ca^{2+} -bound enzyme. TNP-AMP also binds at some distance from a site previously assigned to another ATP analogue, CrATP¹⁵.

Studies of the activation of transport by non-nucleotide phosphates such as acetyl phosphate imply that domain closure to form the phosphorylated enzyme does not require occupation of the nucleotide-binding site. What we need now is a structure for $\text{E}_1\text{Mg}^{2+}\text{P} \cdot (\text{Ca}^{2+})_2$ (Fig. 1) — the phosphoenzyme with Ca^{2+} occluded between the cytoplasm and the lumen. Experiments with glutaraldehyde, which links Lys 492 near TNP-AMP to Arg 678 near the site of phosphorylation of Asp 351 (Fig. 2), have shown that this species is probably the most tightly closed⁷.

In the Ca^{2+} -bound conformation⁴, the A domain is almost dissociated from the main structure, as previously observed¹⁶. This implies that Ca^{2+} , in the absence of nucleotide, loosens interactions between all three domains. To look for possible domain movements, Toyoshima *et al.* rotated the cytoplasmic domains to match the density of the more compact, 8-Å structure of a decavanadate-containing variant of the Ca^{2+} -free conformation of the same pump⁹. The consequent 20° rotation of domain N partially closes the gap between the N and P domains, but still leaves 15–20 Å between the nucleotide and its target — complete closure is prevented by a density ascribed to decavanadate. The 90° rotation required for the A domain brings its highly conserved and mutation-sensitive TGES loop into the site of phosphorylation. The angular relations between helices M4 and M5 and the P domain are also changed.

We can now predict that at least three conformational changes occur during ATP-

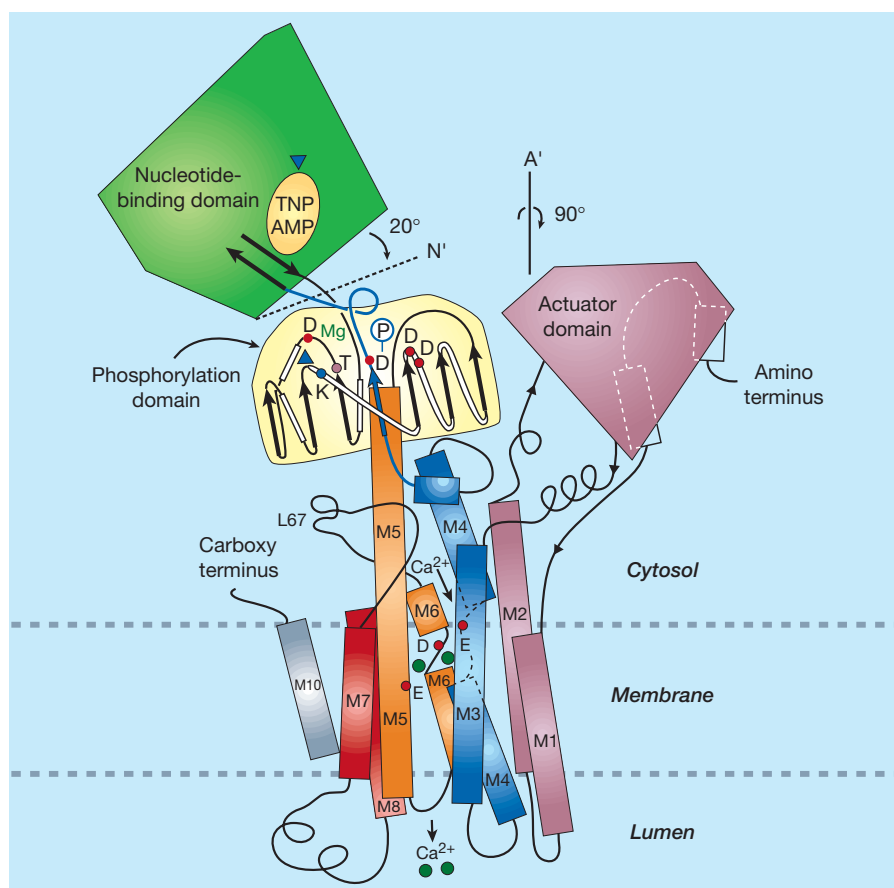


Figure 2 Overview of the structure of the Ca^{2+} -ATPase, as proposed by Toyoshima *et al.*⁴. Small spheres and single letters mark the positions of residues involved in the Ca^{2+} -binding site or in catalysis (K, lysine (blue); T, threonine (violet); D, aspartate, and E, glutamate (red)). N' and A' indicate the rotation of the nucleotide-binding and actuator domains required to match the density of the Ca^{2+} -free conformation⁹. Crosslinked sites are shown as filled triangles and Ca^{2+} ions are shown as green spheres. The membrane region features ten helices, the side-by-side location of the two Ca^{2+} -binding sites formed by transmembrane helices M4, M5, M6 and M8, and the dislocations in M4 and M6 at the common Ca^{2+} -binding motif (E/DGLP). The phosphorylation domain is shown as a seven-stranded parallel β -sheet. The phosphorylated Asp 351 (indicated by a circled 'P') lies at the end of the first strand, which continues into the nucleotide-binding domain.

energized Ca^{2+} transport. The 'jaws' formed by the P and N domains will close to allow phosphorylation of Asp 351. The gap between the A domain and the P and N domains will close⁷. These changes will be transmitted to the Ca^{2+} -binding and translocation domains, first to occlude Ca^{2+} and then to disrupt the Ca^{2+} -binding sites and to alter their accessibility to cytosolic and luminal spaces. Toyoshima *et al.* propose that liberation of the A domain allows it to induce closure of the gap between the N and P domains. At the end of the cycle, the release of Ca^{2+} to the lumen is required before E_2MgP can be hydrolysed (Fig. 1). Perhaps this release provokes contrary motions of the A domain, allowing the cleft to open and admit water to the phosphoenzyme.

The complementary transmission of the effects of phosphorylation to the residues that gate the Ca^{2+} -binding sites is a separate conformational problem^{7,10}. Phosphorylation may initiate small local changes, which may then be communicated from the P

domain to the membrane through the stalk helices and the L67 loop (Fig. 2). There is clearly still more to be learned from this intricate crystal structure of the Ca^{2+} pump. ■

David H. MacLennan is in the Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada.

e-mail: david.maclennan@utoronto.ca

N. Michael Green is at the National Institute of Medical Research, Mill Hill, London NW7 1AA, UK.

e-mail: mgreen@nimr.mrc.ac.uk

1. Ebashi, S. & Lipman, F. J. *Cell Biol.* **14**, 389–400 (1962).
2. Hasselbach, W. & Makinose, M. *Biochem. Z.* **333**, 518–528 (1961).
3. Skou, J. C. *Biochim. Biophys. Acta* **23**, 394–401 (1957).
4. Toyoshima, C. *et al. Nature* **405**, 647–655 (2000).
5. de Meis, L. & Vianna, A. L. *Annu. Rev. Biochem.* **48**, 275–292 (1979).
6. MacLennan, D. H. *et al. Nature* **316**, 696–700 (1985).
7. McIntosh, D. B. *Adv. Mol. Cell Biol.* **23A**, 33–99 (1998).
8. MacLennan, D. H., Clarke, D. M., Loo, T. W. & Skerjanc, I. S. *Acta Physiol. Scand.* (suppl.) **607**, 141–150 (1992).
9. Zhang, P. *Nature* **392**, 835–839 (1998).
10. Stokes, D. L. & Green, N. M. *Biophys. J.* **78**, 1765–1776 (2000).
11. Clarke, D. M. *et al. Nature* **339**, 476–478 (1989).
12. Andersen, J. P. *Biosci. Rep.* **15**, 243–261 (1995).
13. MacLennan, D. H. *et al. J. Biol. Chem.* **272**, 28815–28818 (1997).
14. Soulie, S. *et al. Biochemistry* **38**, 5813–5821 (1999).
15. Yonekura, K. *et al. Biophys. J.* **72**, 997–1005 (1997).
16. Ogawa, H. *et al. Biophys. J.* **75**, 41–52 (1998).

Daedalus

Pumping a vacuum

Last week Daedalus invented a novel aero-engine. Its air duct was threaded by a.c. electric and magnetic fields at right angles. As the electric field came on, it polarized the air, driving a transient displacement current upon which the magnetic field exerted its motor action. As the electric field turned off again, a reversed displacement current flowed through the air; but since the magnetic field had also reversed, the air felt an impulse in the same direction as before. Thus the motor exerted a unidirectional pump action on its gaseous dielectric.

Daedalus now wonders what would happen if he ran this motor in a vacuum. Vacuum is a perfectly good dielectric: it can sustain a transient displacement current like any other insulator. His electromagnetic pump should exert a force on it. So how can you pump a vacuum?

The modern quantum vacuum, he points out, is a very crowded place. It is dense with transient virtual particles appearing and vanishing again all the time, lasting just long enough for their energy not to infringe the uncertainty principle. Clearly it is these particles that carry a displacement current through a vacuum, and on which the electromagnetic vacuum pump exerts its thrust.

Yet any pump or motor must create a reaction equal to the action of its thrust. If the thrust particles promptly vanish, their reaction vanishes with them, which seems impossible. But Daedalus notes the theory of particle collisions: a high-energy collision dumps so much energy in a small volume that the virtual particles in the vicinity can seize it to pay off their debt to Heisenberg, and thus become real. His pump must do the same thing. It dumps energy into the vacuum, allowing real particles to appear. And these newly created particles carry away its reaction.

The fields in the electromagnetic vacuum pump have an extremely low energy density; but their volume as a particle source is enormous. In creating its thrust, the pump would pour out vast numbers of particles of amazingly low energy: languid pions, lethargic neutrinos, feeble photons, all travelling in exactly the direction and with exactly the energy needed to balance its thrust. A splendid new field of physics awaits development. So does a wonderful propellant-free space thruster, which could reach the stars on energy alone.

David Jones

The Further Inventions of Daedalus is published by Oxford University Press.

Recombinant erythropoietin in urine

An artificial hormone taken to boost athletic performance can now be detected.

Erythropoietin is a hormone that stimulates the production of new red blood cells (erythropoiesis). Although athletes use recombinant human erythropoietin illicitly to boost the delivery of oxygen to the tissues and enhance their performance in endurance sports, this widespread doping practice cannot be controlled in the absence of a reliable analytical procedure to monitor it. Here we describe a new technique for detecting this drug in urine following its recent administration.

The stimulation of erythropoiesis by erythropoietin (EPO) makes this drug very attractive to sportspeople wishing to improve their aerobic power, although the International Olympic Committee banned its misuse ten years ago. Detection has been a problem — analysis of haematological¹ or biochemical² parameters indicates only that erythropoiesis has been stimulated, but cannot confirm that drug administration is to blame.

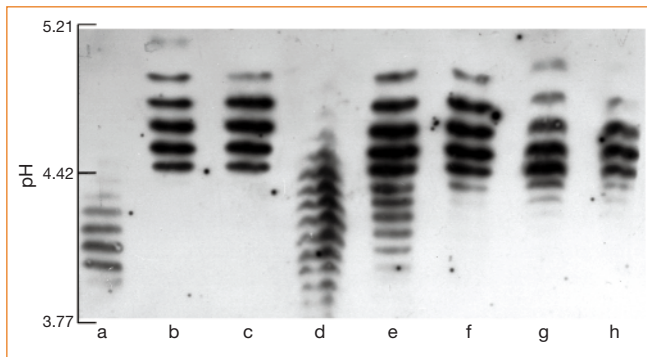
To detect administered hormone directly means that exogenous, recombinant EPO must be differentiated from natural, endogenous EPO. A promising electrophoretic method³ has proved impractical for screening by the antidoping laboratories. We have developed an analytical procedure for detecting recombinant EPO in urine and have applied it to specimens from cyclists participating in the the infamous Tour de France 1998 competition, which was sullied by scandals about EPO doping.

Owing to microheterogeneity in their structures, natural and recombinant EPO comprise several isoforms, some of which have charge differences and can be separated by isoelectric focusing (Fig. 1). We found that the isoelectric patterns of the two recombinant EPO- α and - β forms are very similar (both have an isoelectric point, pI, in the range 4.42–5.11); although EPO- β has an extra basic band, both differ from natural, purified urinary EPO, which has more acidic bands (pI 3.92–4.42), probably due to post-translational modifications such as glycosylation, which is species- and tissue-type-dependent⁴. Such differences in the urine analysis allowed us to ascribe excreted EPO to a natural or recombinant origin.

We developed an immunoblotting procedure to obtain a reliable image of EPO patterns in urine. Our results (Fig. 1) indicated that the patterns from control subjects consisted of about 10 bands of pI 3.77–4.70, in accord with the purified natural urinary EPO pattern, whereas those from subjects treated with recombinant EPO contained more basic bands, reflecting

Figure 1 Autoradiograph of isoelectric patterns of exogenous and endogenous erythropoietin (EPO). Images were obtained by chemiluminescent immunodetection of blotted EPO after isoelectric focusing. **a**, Purified commercial human urinary natural EPO (Sigma); **b**, recombinant EPO- β (Neorecomon, France); **c**, recombinant EPO- α (Eprex, France); **d**, urine from a control subject; **e,f**, urine from two patients treated with

Neorecomon EPO for post-haemorrhagic anaemia; **g,h**, urine from two cyclists from Tour de France 1998 (samples concentrated by ultrafiltration). Note the 'mixed' appearance of the pattern in **e**. The cathode is at the top; pH values are indicated on the left.



the presence of recombinant isoforms, and sometimes acidic bands as well, depending on the presence of endogenous isoforms. The presence of exogenous hormone was always evident: any individual injected with recombinant EPO showed a striking transformation of their initial EPO urine pattern.

We assayed 102 frozen urine samples from participants in the Tour de France 1998 cycling competition for EPO by using an enzyme-linked immunosorbent assay. Twenty-eight of these samples had EPO levels above the normal range of 0–3.7 international units per litre (mean, 0.48 IU per litre, $n = 103$; 77 samples were below the minimum detectable concentration of 0.6 IU per litre). We analysed the 14 samples presenting with the highest concentrations (7–20 IU per litre): although characterization of the EPO source does not require such high levels for urine analysis, we

selected these samples for isoelectric focusing as they were more likely to contain exogenous hormone; indeed, they all gave rise to a banding pattern typical of recombinant hormone.

Our method for detecting recent exposure to recombinant EPO in athletes could be useful for in-competition controls in events of long duration (for example, cyclists have been known to use exogenous EPO continuously for 6 months at a time), but should find its principal application in out-of-competition testing.

Françoise Lasne, Jacques de Ceaurriz

National Anti-Doping Laboratory,
92290 Châtenay-Malabry, France

1. Casoni, I. *et al. Int. J. Sports Med.* **14**, 307–311 (1993).
2. Gareau, R. *et al. Nature* **380**, 113 (1996).
3. Wide, L. *et al. Med. Sci. Sports Exerc.* **27**, 1569–1576 (1995).
4. Rademacher, T. W., Parekh, R. B. & Dwek, R. A. *Annu. Rev. Biochem.* **57**, 785–838 (1988).

Phylogeny

Parabasal flagellates are ancient eukaryotes

Discrepancies between eukaryotic phylogenetic trees based on different gene sequences have led to the suggestion that the deepest branches of each gene tree could simply be artefacts of rapid evolution rather than indicators of an ancient divergence^{1–3}. But if an insertion or deletion occurred in a gene sequence very early in eukaryotic evolution, the oldest eukaryotic lineages should be recognizable by their resemblance to prokaryotes lacking this character. Here we investigate the structure of the gene encoding enolase, an enzyme of the glycolytic pathway, and find that the gene from parabasal flagellates lacks two deletions present in other eukaryotic

enolases, indicating that Parabasalia could be the most ancient eukaryotes examined so far.

Eukaryotic enolase sequences contain several insertions and deletions compared with each other, Eubacteria and Archaeobacteria, some of which have been used to link animals and fungi⁶. We sequenced enolase genes from three putatively ancient lineages: diplomonads, Parabasalia and kinetoplastids. Neither kinetoplastid nor diplomonad enolase genes are exceptional (nor is that of *Entamoeba* enolase, another putatively ancient eukaryote), but parabasal enolases lack two close, single-amino-acid deletions common to all other eukaryotic enolases (Fig. 1a, overleaf).

Given the proximity of these deletions, they may have resulted from a single event. However, the surrounding alignment is reproducible, and the amino acids at these

positions are identical or similar in Parabasalia and prokaryotes. The simplest explanation is that parabasal enolases have retained the ancestral, prokaryotic state while the common ancestor of other eukaryotic enolases sustained two deletions. This is consistent with enolase phylogeny (Fig. 1b), although the position of parabasal enolases must be treated with the extreme caution demanded by deep branches^{1–5}, particularly as parabasal enolases are quite divergent. The phylogeny is important, however, because it shows that these deletions in the enolase gene do not simply reflect a lateral transfer of enolase from a prokaryote to Parabasalia.

Other explanations for these observations cannot be excluded, such as parallel deletions in two or more lineages of non-parabasalian eukaryotes, deletions in the common ancestor of all eukaryotes and subsequent re-insertion of these two amino

acids in Parabasalia, or recombination with a prokaryotic enolase. However, these are all more complex, and the consistency between enolase deletions and molecular phylogeny, and the lack of opposing evidence, favours the explanation that Parabasalia are the earliest extant lineage of eukaryotes studied.

Clues to the nature of the ancestral eukaryote might be gleaned by comparing Parabasalia with other eukaryotes. One difference is that Parabasalia lack mitochondria, containing instead an anaerobic metabolic organelle called a 'hydrogenosome'⁷. The parabasal hydrogenosome is derived from the mitochondrial lineage^{7,8}, and hydrogenosomes have evolved from mitochondria in several other lineages⁸. Parabasal hydrogenosomes may also have evolved from a fully specialized mitochondrion, but, if Parabasalia are extremely ancient, their hydrogenosomes might have evolved from an unspecialized proto-

mitochondrial symbiont that developed its bioenergetic capacity by a different route from typical mitochondria^{7,9}.

A better understanding of this and other atypical features of Parabasalia (for example, mitosis, meiosis and transcription initiation¹⁰) may reveal whether such characters reflect the primitive state of eukaryotes or odd derivations in Parabasalia. Conversely, the ordinary aspects of Parabasalia (they too have a nucleus, cytoskeleton, flagella, Golgi bodies and spliceosomes¹¹) indicate that many important characteristics were established in the ancestor of extant eukaryotes. Parabasalia, then, are not major intermediates in a gradual transition between prokaryote and eukaryote and, should such an intermediate lineage still exist, it is unlikely to have been characterized.

Discovering primitive molecular characteristics is perhaps our best hope of identifying ancient eukaryotic lineages, although

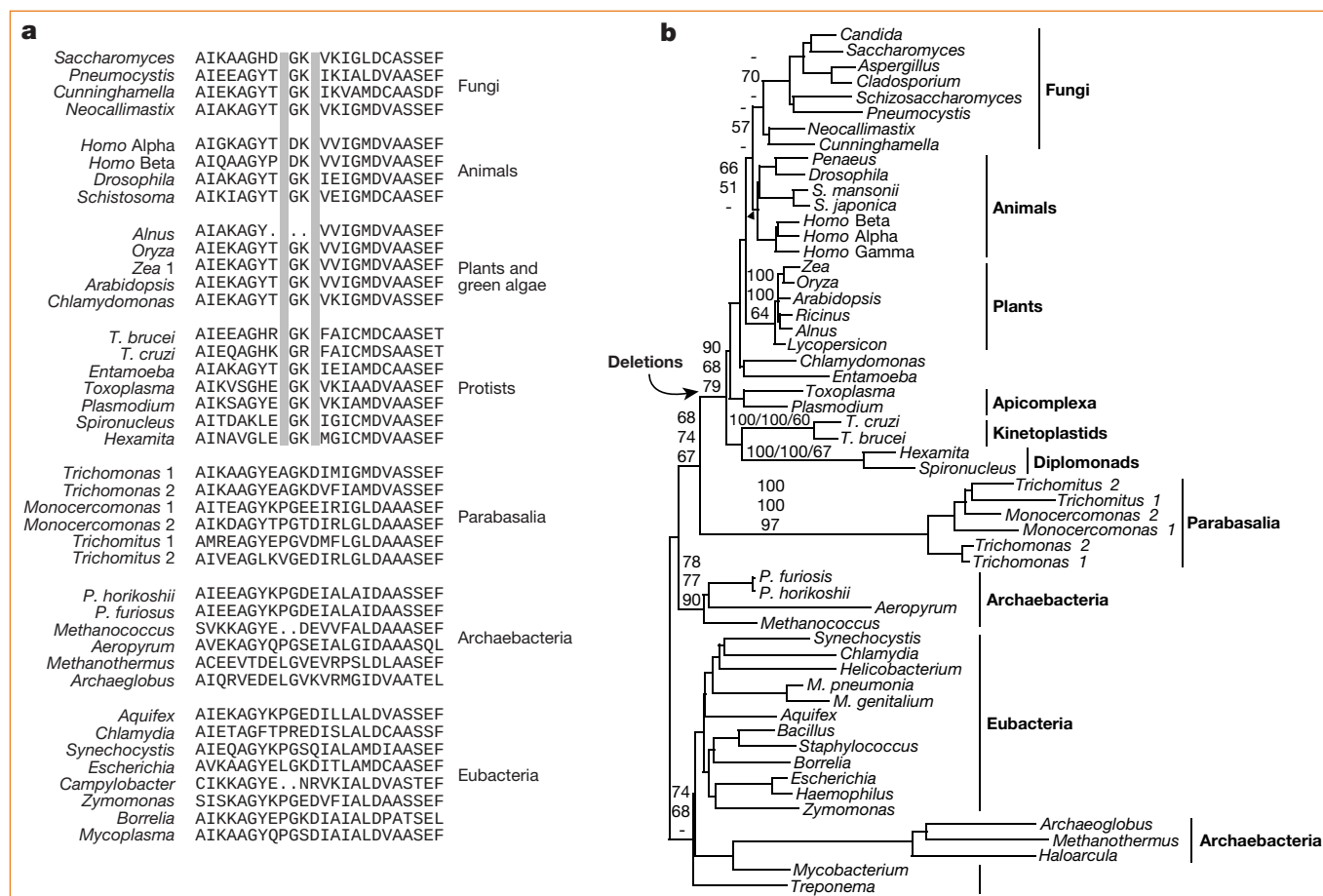


Figure 1 Evidence from enolase structure that Parabasalia are the most ancient eukaryotes. **a**, Deletions in enolase found in all eukaryotes except Parabasalia (shaded). Independent deletions also occurred in *Alnus*, *Methanococcus* and *Campylobacter* (not shaded). Positions shown correspond to amino acids 232–253 of the *Saccharomyces* gene. **b**, Phylogeny of enolase. Numbers at nodes are bootstrap values without gamma correction (top), bootstrap values with gamma correction (middle), and per cent occurrence in 10,000 quartet puzzling trees (bottom). Only values over 50% for major nodes are shown. Enolase genes were amplified from *Spironucleus vourmens* (ATCC 50386), *Hexamita inflata* (AZ-4), *Trypanosoma cruzi* (Cl-Brenner), *T. brucei* (IsTat 1.1a), *Trichomonas vaginalis* (NIH-C1, ATCC 30001), *Trichomonas batrachorum* (G11), and *Monocercomonas* sp. (Ns-1PRR, ATCC 50210). DNA was provided by M. Müller, H. van Keulen and J. Feagan. Polymerase chain reaction products were cloned using TOPO TA vector (Invitrogen) and multiple clones from each species were sequenced on both strands. GenBank accession numbers for new sequences are AF159517–AF159532. Inferred amino-acid sequences were aligned with known homologues using PIMA and edited manually. The region shown in **a** was realigned by different investigators using several algorithms and models, consistently giving the alignment shown. Trees were inferred from 372 amino acids deemed to be homologous among all three domains. Distances and bootstrap distances (using Puzzleboot, by M. Holder and A. Roger) were calculated with PUZZLE 4.0.1 using the JTT substitution matrix and amino-acid frequency estimated from the data. Site-to-site rate variation was modelled on a gamma distribution with eight rate categories plus invariant sites and the shape parameter was estimated from the data. Trees were constructed using BioNJ (other methods also gave the same placement of Parabasalia). Maximum-likelihood trees were inferred using quartet/9 puzzling with 10,000 puzzling steps and the PUZZLE settings described above.

sadly our sampling from those protists most likely to contain these is limited. Other characteristics distinguishing Parabasalia as the earliest eukaryotic lineage are waiting to be discovered — confirmation of the ancient origin of Parabasalia may depend on these characteristics being common in parabasalian genomes.

Patrick J. Keeling*†, Jeffrey D. Palmer*

*Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

†Present address: Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

e-mail: pkeeling@interchange.ubc.ca

- Embley, T. M. & Hirt, R. P. *Curr. Opin. Genet. Dev.* **8**, 624–629 (1998).
- Philippe, H. & Laurent, J. *Curr. Opin. Genet. Dev.* **8**, 616–623 (1998).
- Keeling, P. J. *BioEssays* **20**, 87–95 (1998).
- Stiller, J. W. & Hall, B. D. *Mol. Biol. Evol.* **16**, 1270–1279 (1999).
- Roger, A. J. *et al. Mol. Biol. Evol.* **16**, 218–233 (1999).
- Baldauf, S. L. & Palmer, J. D. *Proc. Natl Acad. Sci. USA* **90**, 11558–11562 (1993).
- Müller, M. *Parasitol. Today* **13**, 166–167 (1997).
- Embley, T. M., Horner, D. A. & Hirt, R. P. *Trends Ecol. Evol.* **12**, 437–441 (1997).
- Martin, W. & Müller, M. *Nature* **392**, 37–41 (1998).
- Liston, D. R. & Johnson, P. J. *Mol. Cell. Biol.* **19**, 2380–2388 (1999).
- Fast, N. M. & Doolittle, W. F. *Mol. Biochem. Parasitol.* **99**, 275–278 (1999).

Planetary science

Tectonics and water on Europa

The tectonics of Europa, one of Jupiter's moons, are complex. This satellite probably hosts a subsurface water ocean, but the thickness of the outer ice crust is poorly constrained and the episodic presence of liquid water at the surface is debated. We argue that some surface features of Europa are formed by soft ice that is heated by viscous dissipation of tidal motion along faults, and do not depend on a shallow ocean. Our model suggests that transient pockets of liquid water or brine could form at shallow depths in the crust.

The Galileo spacecraft has returned high-resolution images of surface lineations and ridges on Europa. These have prompted proposals that tidal stresses from Jupiter may fracture the icy crust of the synchronously rotating satellite, allowing material from the interior to reach the surface^{1,2}. One model invokes tidal expansion and contraction of metre-wide cracks which repeatedly force liquid water up from a subsurface ocean to the surface of the moon as it completes each 85-hour orbit³ — a process that depends on a thin (1 km or less) ice shell and which has fuelled speculation that the ocean frequently meets the sunlit surface and could be a habitat for life.

The formation of kilometre-deep cracks

in Europa's crust is a problem. Below a depth of 35 m, the pressure from the weight of the overlying ice will exceed the estimated stresses due to tides ($< 4 \times 10^4$ Pa) and prevent cracks from growing. Any fractures will halt at a depth where warmer, less viscous ice flows, rather than fracturing further, to accommodate tidal strain. Taking an estimated strain rate of 2×10^{-10} s⁻¹ and a relation between normal stress and yield stress appropriate for ice⁴, we find that the brittle–ductile transition occurs at depths where temperatures⁵ exceed 170 K, well above those of any ice–ocean interface.

Also, liquid water in a crack will freeze solid by virtue of the conduction of heat to the walls: the freezing time is $t = w^2 / (16\kappa\lambda^2)$, where w is the crack width, κ is the thermal diffusivity of ice (1.7×10^{-6} m² s⁻¹) and λ is a dimensionless parameter⁶ equal to about 0.3. A 1-m crack will freeze in 1.3 orbits. Hydrostatic forces are too weak to extrude ice from such a narrow orifice and, unless a compensatory removal of crust occurs elsewhere, the crack will disappear.

We propose, alternatively, that tides drive viscous flow and heating by dissipation at zones of lateral motion (strike–slip) in the crust. There is accumulating evidence for strike–slip motion on Europa⁷. The relative motion along a fault or defect will produce frictional heating, causing the local temperature to increase and the viscosity of the ice to decrease⁸. This feedback can lead to accommodation of the relative motion of two blocks of crust by viscous flow in a zone of finite width, rather than a discontinuity at a fault. Steady-state conditions in the zone are derived by equating the production of heat by viscous dissipation with its loss by conduction to the surrounding ice.

Strike–slip motion of amplitude a at a diurnal period t (3×10^5 s) will maintain a temperature T_c and viscosity η_c at the centre of the shear zone⁶ that satisfy $16kRT_c^2t^2 = \pi^2\eta_cE_a a^2$, where E_a and k are the activation energy (6×10^4 J at 273 K) and thermal conductivity (2.1 W m⁻¹ K⁻¹ at 273 K), respectively, of the ice, and R is the gas constant.

A plausible diurnal motion³ of 0.6 m can maintain the ice in the centre of the shear zone at a temperature of 273 K, where its viscosity will be roughly 10^{13} Pa s. The width of this zone of soft ice is $\delta = \eta a(\tau t)^{-1}$, where τ is the shear stress (about 1 km for $\tau = 2 \times 10^4$ Pa). This warm ice will have a buoyant density contrast $\Delta\rho = 20$ kg m⁻³ with respect to the surrounding ice and will flow upwards by a few tens of centimetres over the course of one tidal cycle.

We suggest that such motion over the course of many cycles could be responsible for the formation of structures such as ridge pairs. Our model does require the existence of an ocean (not necessarily shallow): without the mechanical decoupling between the

ice crust and the interior, the motion of the crust would be 30–50 times smaller.

Larger strike–slip motion may lead to partial melting (liquid water) in the shear zone. The melt-generation rate scales as $\tau^2(\rho\eta_c L)^{-1}$, where L and ρ are the latent heat of fusion (3.25×10^5 J kg⁻¹) and density (900 kg m⁻³), respectively, of ice. This production will be balanced by downward percolation of the denser melt at a rate⁹ $A\phi^n\Delta\rho g(\eta_m h)^{-1}$, where A and n are constants, ϕ is the melt fraction, η_m is the melt viscosity, h is the thickness of the melt column, and g (1.3 m s⁻²) is Europa's surface gravity. Although there are considerable uncertainties in some of these values, the equilibrium melt fraction is of the order of 1%. This much melt will form only if the strike–slip motion is sufficiently large (~ 1 m) to compensate for the reduction in ice viscosity by a factor of about one third as a result of the presence of melt¹⁰.

The pore pressure due to the presence of melt will also increase the depth to the brittle–ductile transition and allow fractures to accommodate strike–slip motion and to slow heat generation. Melt pockets below the fracture zone will percolate downwards at a velocity of a few tens of metres per year, and so will have a lifetime of $\sim 1,000$ years. Any melt at the base of the fracture zone will escape through the fractures or freeze and, if salts are present, brines will be rejected from the freezing melt¹¹.

In this case, the melt lifetime, estimated by equating the latent heat that must be rejected to the rate of thermal conduction away from the fault zone, will be about 30 years (although shorter at shallow depths, where vertical conduction to the surface is important). Depending on the thickness of the fractured zone, transient liquid-water or brine pockets may exist within reach of sunlight, potentially providing habitats for photosynthetic organisms capable of remaining dormant in ice for millennia¹² between relatively brief 'blooms'.

Eric J. Gaidos*, Francis Nimmo†

*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA

†Bullard Laboratories, Department of Earth Sciences, Madingley Rise, Madingley Road, University of Cambridge, Cambridge CB3 0EZ, UK

- Sullivan, R. *et al. Nature* **391**, 371–373 (1998).
- Pappalardo, R. T. *et al. J. Geophys. Res.* **104**, 24015–24055 (1999).
- Greenberg, R. *et al. Icarus* **135**, 64–78 (1998).
- Rist, M. A. *J. Phys. Chem. B* **101**, 6263–6266 (1997).
- Pappalardo, R. T. *et al. Nature* **391**, 365–367 (1998).
- Turcotte, D. L. & Schubert, G. *Geodynamics: Applications of Continuum Physics to Geological Problems* (Wiley, New York, 1982).
- Tufts, B. R. *et al. Icarus* **141**, 53–64 (1999).
- Stevenson, D. J. in *Europa Ocean Conf., Capistrano Conf.* 5 69–70 (San Juan Capistrano Res. Inst., CA, 1996).
- McKenzie, D. *Earth Planet. Sci. Lett.* **74**, 81–91 (1985).
- Duval, P. *Int. Assoc. Sci. Hydro.* **118**, 29–33 (1977).
- Cottier, E., Eiken, H. & Wadhams, P. *J. Geophys. Res.* **104**, 15850–15871 (1999).
- Vorobyova, E. *et al. FEMS Microbiol. Rev.* **20**, 277–290 (1997).

Scalar turbulence

Boris I. Shraiman* & Eric D. Siggia†

* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Center for Physics and Biology, Rockefeller University, New York, New York 10021, USA

The advection of a passive substance by a turbulent flow is important in many natural and engineering settings. The concentration of such a substance can exhibit complex dynamic behaviour that shows many phenomenological parallels with the behaviour of the turbulent velocity field. Yet the statistical properties of this so-called ‘passive scalar’ turbulence are decoupled from those of the underlying velocity field. Passive scalar turbulence has recently yielded to mathematical analysis, and such progress may ultimately lead to a better understanding of the still intractable problem of fluid turbulence itself.

The concentration of a substance advected by a turbulent flow exhibits a complex, chaotically evolving structure over a broad range of space and time scales. The “substance” could be a pollutant, as in the familiar case of smoke dispersing in air; it could be heat, when a hot object is cooled in the flow; or (as in Fig. 1) a fluorescent dye mixed by a turbulent jet. Turbulent advection is important in many natural and engineering settings, ranging from atmospheric phenomena¹ and combustion² to the stretching and amplification of magnetic fields on both galactic and terrestrial scales³. It is even relevant in a biological context. For example, many organisms need to locate the sources of attractive odours: the chemotactic ‘algorithms’ by which this is achieved will differ according to whether the organism is in a diffusion-dominated world (as in the case of bacteria⁴) where local concentration gradients can be used to navigate, or in a turbulent world (like lobsters⁵ or moths⁶) where large flow-induced fluctuations of the local concentration gradient render it useless for navigation. In many cases, the advected “substance” has a strong effect on the turbulent flow itself, by generating local forces: for example, non-uniform heating would act on the flow through buoyancy. However, the phenomenon of turbulent advection and mixing caused by it may be separated from the forces responsible for the flow.

Here we concentrate on the case where the advected substance is passive and so has a negligible back effect on the flow (this also applies to the temperature field if buoyancy forces are small compared to the inertial stresses driving the flow), and on situations of well controlled laboratory turbulence^{7,8}. Given that the substances are passive and are described by scalar (rather than vector) fields, they all behave in the same way. Thus generically we speak of ‘passive scalar’ advection.

Turbulent flow transports and disperses the scalar by making parcels of fluid follow chaotic trajectories. The spatial non-uniformity of the velocity field causes the lines of constant scalar concentration to stretch and fold; as a result, variations of the scalar concentration reach progressively smaller scales. This process amplifies the local concentration gradients until molecular diffusivity finally takes over, causing local variations of the scalar to dissipate. The result of these processes is a dramatic acceleration of the rate of mixing which, for a fully turbulent flow, becomes independent of molecular diffusivity. However, the price of the (on average) rapid mixing is the frequent occurrence of large fluctuations of the scalar field, which are quantified by departures of the probability distribution functions (PDF) of the scalar field from gaussian behaviour^{8,9}. Turbulent mixing, defined as the dissipation of scalar variance, is “intermittent”^{7,9,10}—it occurs in spatio-temporal bursts qualitatively similar to the energy dissipation by the turbulent velocity field itself. Although the dynamics of the underlying velocity (vector) field is intrinsically nonlinear and much more complex to describe, there are so many parallels between the statistics of the scalar fluctuations and those of

turbulent velocity that one may justifiably refer to the behaviour of the advected scalar as ‘scalar turbulence’. The basic understanding of turbulent mixing and transport emerged through the work of Taylor, Richardson, Kolmogorov, Obukhov and Corrsin (as reviewed by Monin and Yaglom¹¹). However, while their simple dimensional arguments^{11,12} are successful in explaining the average rate of spreading or mixing of, say, a smoke plume, understanding the statistics of the large fluctuations is a more difficult problem. We will address this below.

The practical relevance of understanding the statistics of large fluctuations is obvious if one considers, for example, the probability of a pollutant concentration exceeding some tolerable level as it spreads from a source. More subtle, but equally important, is the role of large concentration fluctuations in controlling the rate of slow (high order) chemical reactions, for example in the process of atmospheric ozone destruction¹³.

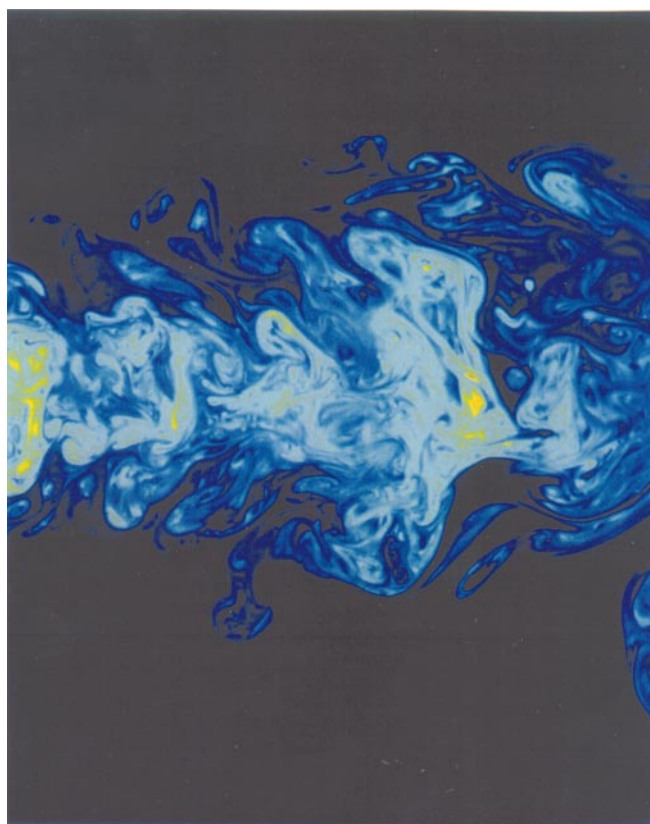


Figure 1 Fluorescent dye in a turbulent jet. The Reynolds number, Re , is 4,000 (K. R. Sreenivasan).

This Review is prompted by recent progress in the statistical description of passive scalar turbulence. Central to this progress has been the realization that anomalous scaling properties and the appearance of coherent structures in the scalar field—well known characteristics of fluid turbulence itself—occur even for a scalar advected by a simple random gaussian velocity field (which resembles real turbulence only in the way the fluctuations are distributed over small spatial scales^{14–17}). The non-trivial statistical aspects of the scalar turn out to originate in the mixing process itself, rather than being inherited from the complexity of the turbulent velocity field. Study of passive scalar turbulence is therefore decoupled from the still intractable problem of calculating the velocity statistics, and so has yielded to mathematical analysis^{18–20}. Conversely, the well established phenomenological parallels between the statistical description of mixing and fluid turbulence itself^{9,10} suggest that progress on the latter front may follow from a better understanding of turbulent mixing.

Observations and phenomenology

The simplest characterization of turbulent mixing is through a local measurement of the advected field. Measurements of temperature at a given point in a turbulent flow—for example, on the axis of a hot jet mixing with a cold bath²¹ or within a turbulent boundary layer over a heated surface^{22,23}—are typical of many laboratory experiments. Closer to the idealized case of ‘homogeneous and isotropic’ turbulence are the recent studies by Warhaft and co-workers^{24,25}, in which temperature fluctuations were measured behind the grid in a wind tunnel where the inlet flow was preheated to impose a linear temperature profile (in a direction transverse to the flow). In all of these cases, the temperature measured downstream from the inlet as a function of time, $\Theta(t)$, fluctuates about its mean value $\langle\Theta(r)\rangle$ at the point of observation r as parcels of colder or warmer fluid sweep by. The fluctuations are then quantified by constructing, from the temporal sequence, a histogram or PDF, $P(\Theta)$. If the scalar field is injected randomly¹¹, the resulting single-point PDF is gaussian. In the more common case when the ‘injection’ produces a large-scale temperature gradient^{24,26}, a regime of exponential tails (see Fig. 2a) is observed. In either case, the root-mean-square (r.m.s.) fluctuation is of the same order of magnitude as the temperature variation on the scale characteristic of the flow as a whole—the so-called ‘integral scale’, which is set by the geometry (for example, the jet diameter or wind tunnel grid).

Another important observable is the difference between simultaneous measurements at two points separated by a distance r , expressed as $\Delta_r\Theta \equiv \Theta(x+r, t) - \Theta(x, t)$. The PDF of the difference is shown in Fig. 2b: it starts as a gaussian for point separations of the order of the integral scale L , and evolves progressively longer exponential or stretched-exponential tails as r decreases towards the dissipation scale η , where the scalar difference can be approximated by a derivative. This excess of large fluctuations (compared to a gaussian distribution) on small scales, termed ‘intermittency’, is well known for the statistics of the velocity fluctuations themselves ($\Delta_r v$)^{9,10}.

The behaviour of the scalar field is determined by two physical effects: (1) transport, the physical translocation of impurities or heat by the combined action of the flow and diffusion; and (2) mixing, the irreversible decay of fluctuations, ultimately due to the molecular diffusivity, κ , which (in the absence of forcing) will tend to reduce the scalar field to uniformity. In turbulent flow, both the transport and mixing rate become independent of κ when the diffusivity is infinitesimally small. (The latter limit is highly non-trivial because diffusivity, no matter how small, dominates over advection by the flow at sufficiently small scales and expresses the irreversibility of molecular mixing.)

Taylor proposed¹² that the motion of a single particle in turbulent flow is diffusive, and given by $\langle R^2(t) \rangle \sim D_{\text{eff}} t$ (where $R(t)$ is the particle displacement in the reference frame moving with the mean

flow velocity) with the effective diffusivity of the order of $D_{\text{eff}} \approx \Delta V^2 \tau_L$ (where ΔV is the r.m.s. velocity fluctuation at a point, and τ_L is the velocity correlation time). As the correlation length of the velocity field is the integral scale L , the correlation time τ_L is estimated dimensionally as $L/\Delta V$, and the effective diffusivity D_{eff} turns out to be independent of κ .

Single-particle diffusion must be contrasted with the dispersion of nearby particles, which also governs the spreading of the ‘plume’ that appears when the scalar is injected locally. According to Richardson¹², the relative motion of a pair of particles with separation $r < L$ is governed by $\langle r^2(t) \rangle \sim \epsilon t^3$, where $\epsilon \approx \Delta V^2/\tau_L$ is the energy dissipation rate for the turbulent velocity field^{10,12}. Richardson’s law only holds for short times (while $r \leq L$), but does predict that two particles will separate to the integral scale in a time independent of their initial separation. Although according to Richardson r^2 grows as a power of t rather than linearly, in fact the particles are separating more slowly than they would under Taylor diffusion (note that $\epsilon \sim D_{\text{eff}}/\tau_L^2$). This happens because the motion of particles within an integral scale of each other is correlated. For $\tau > \tau_L$ and $r > L$, particles decorrelate and Richardson’s law crosses over to Taylor’s law. In fact, Richardson’s law is obtained by assuming relative motion to be diffusive but with a diffusion constant $D(r) \sim \epsilon^{2/3} r^{4/3}$ which increases with r . This behaviour follows from Taylor’s argument once one takes into account that the variance of the (relative) velocity fluctuations on scale r is $\langle (\Delta_r v)^2 \rangle \sim \epsilon^{2/3} r^{4/3}$ and the corresponding correlation time is $\tau_r \approx r/\Delta_r v \sim \epsilon^{-1/3} r^{2/3}$. The last two expressions are the predictions of Kolmogorov’s 1941 (K41) scaling theory¹⁰.

Kolmogorov’s theory¹⁰ postulates the following: (1) the energy dissipation rate ϵ is independent of the viscosity ν provided that the Reynolds number $\text{Re} = \Delta V L/\nu$ is sufficiently large (that is, when

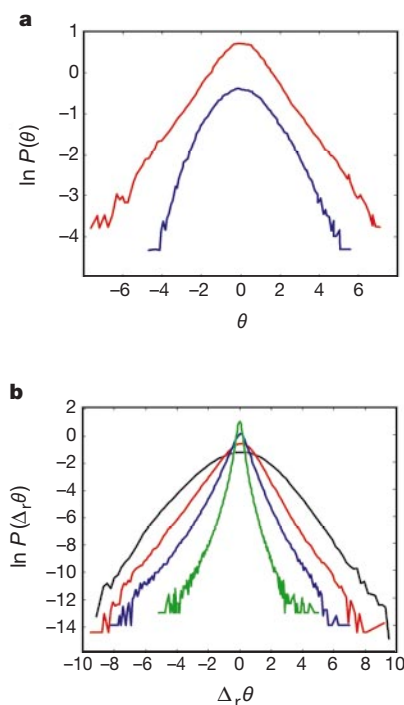


Figure 2 Probability distribution functions (PDF). **a**, $P(\theta)$ 62 grid lengths downstream in a wind tunnel with (blue) and without (red) a mean gradient (from ref. 24); the scalar field θ is scaled to unity for each curve. **b**, The PDF of scalar difference $\Delta_r\theta$ at different separations starting from the integral scale $r = 230\eta$ (black), through to the inertial range $r = 96\eta$ (red), 26η (blue) and 6.5η (green) (from F. Moisy, W. Herve and P. Tabeling, manuscript in preparation). Note that the r.m.s. decreases with decreasing scale but the fluctuations that are large compared to the r.m.s. become more likely as the PDFs evolve from gaussian on the integral scale to a stretch-exponential on small scales.

inertial forces dominate over viscous forces); (2) ϵ is the *only* relevant parameter which controls turbulent fluctuations on scales $r \ll L$. The first postulate has strong experimental support, as does the aforementioned 2/3 law for the velocity variance as a function of r (refs 9–12). Kolmogorov's argument was extended to passive scalar by Obukhov and Corrsin¹¹. The Kolmogorov–Obukhov–Corrsin theory states that the dissipation rate of scalar variance^{11,12}, $\epsilon_\theta \equiv -d/dt\langle(\Delta\theta)^2\rangle$, is independent of molecular diffusivity and is set by the large eddy turnover rate τ_L^{-1} , so that $\epsilon_\theta \sim \langle(\Delta\theta)^2\rangle/\tau_L$. Because dissipation occurs only on small scales, the same scalar variance flux must pass through the ‘inertial range’ comprising all scales intermediate between the initial injection scale L and the final dissipation scale η (see below). Using the K41 estimate for $\Delta_r v$ on scale r , one obtains $\langle(\Delta_r\theta)^2\rangle \sim \epsilon_\theta \epsilon^{-1/3} r^{2/3}$. Molecular diffusivity enters only in setting the scalar dissipation scale η by matching the dissipation rate defined by $\kappa\langle(\Delta_\eta\theta/\eta)^2\rangle \sim \epsilon_\theta$.

Kolmogorov–Obukhov–Corrsin (KOC) theory also predicts simple scaling for the time average of higher powers of $\Delta_r\theta$, (the moments of the PDF) or the so-called structure function, $S_{2n}(r) \equiv \langle(\Delta_r\theta)^{2n}\rangle \approx \langle(\Delta_r\theta)^2\rangle^n$. Therefore a universal PDF of scalar differences is predicted, $P(\Delta_r\theta/\langle(\Delta_r\theta)^2\rangle^{1/2})$, where all the dependence on the measurement scale is accounted for by normalizing to the r.m.s. of the fluctuations. Although the KOC prediction for the second moment is close to experimental values^{7,8}, its predictions for the high moments are far off: as evident from the strong dependence of the shape of the difference PDF on scale r , Fig. 2b. The non-Kolmogorov behaviour of large fluctuations can be quantified by the observed^{8,9} anomalous scaling of the moments: $S_{2n}(r)/S_2^n(r) \approx (r/L)^{\zeta_{2n}-2n/3}$. The deviation of the scaling exponents ζ_{2n} from their

KOC values ($2n/3$) are even larger than those known for the velocity structure function^{9,10}, indicating that the intermittency in the scalar field is even stronger. The anomalous scaling means that ϵ_θ (like ϵ in the K41 theory) is not the only relevant parameter for $r < L$. On purely dimensional grounds, the integral scale L must enter inertial range quantities. However, there is no unique definition of L and numerically any expression for L will reflect the anisotropy of the large scales. Is it even possible that the small scales are anisotropic, imprinted by the anisotropies on the large scales, and therefore quite non-universal? The definitive answer is still not known, yet the study of the passive scalar problem sheds light on the mechanism that could apply to the velocity as well.

Before returning to the failure of the K41 theory in describing the fluctuations, we will look further into the physics underlying the remarkable and correct first postulate: the independence of the dissipation rate on the viscosity and molecular diffusivity κ . Consider first a passive scalar in a random flow characterized by single length and time scales, L and $\tau_L \approx \Delta V/L$, respectively. The characteristic magnitude of the velocity gradient in this so-called Batchelor flow regime is $\gamma = \Delta V/L$. A scalar blob will be stretched and rolled up by the flow and, because of the incompressibility, acquire structure—such as tendrils and ‘Swiss roll’ shapes—on progressively smaller scales $l(t) \approx L \exp(-\gamma t)$. Molecular mixing occurs rapidly and gradients disappear at the time t_* , when the folds of the scalar field reach down to the scale where the diffusion rate becomes comparable to the rate of stretching, $\kappa/l^2(t_*) \approx \gamma$. This defines the dissipation scale for Batchelor flow expressed as $\eta_B \approx L/\text{Pe}^{1/2}$, which depends on κ through the Peclet number $\text{Pe} \equiv \Delta V L/\kappa \gg 1$ and the mixing time $t_* \approx \gamma^{-1} \ln \text{Pe}$ which is

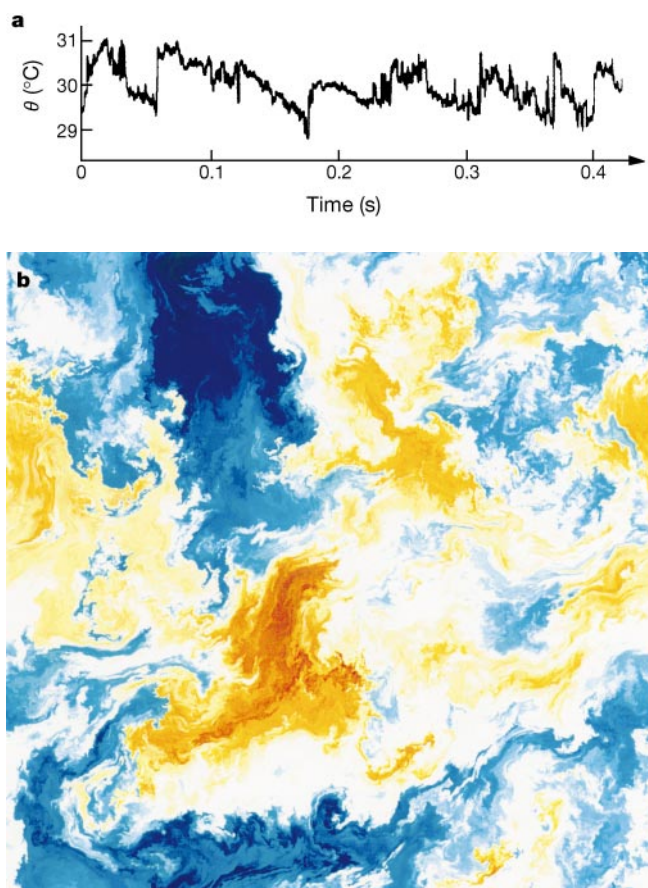


Figure 3 Scalar fluctuations in time and space. **a**, Temporal trace of the temperature recorded at a fixed point in a turbulent boundary layer over a heated plate from ref. 23. Note the asymmetry of the derivative. **b**, Numerical simulation of passive scalar advection

in two dimensions for the Kraichnan model with velocity defined by equation (2) and $\gamma = 0.5$ on a 8192 square grid³¹. The concentration scale runs from blue to red.

only weakly (logarithmically) dependent on κ . The rate of mixing and scalar dissipation follows t_*^{-1} . (This argument tacitly assumes the flow to be sufficiently random not to be concerned with the possible existence of stagnation points, separatrices and subregions left invariant by the flow. Their appearance is possible in weakly time-dependent (single-scale) flows even when trajectories on average separate exponentially. The latter is the case for chaotic advection which is an important instance of mixing^{27,28} in its own right. Realization of Batchelor's random flow limit has been plausibly achieved in several experiments^{29,30}, although observation of Batchelor's scaling¹² remains elusive.)

In contrast to the Batchelor regime, flows at high Re number involve not one but many length scales. But the above argument can be used on each octave of scales (starting with the dissipation scale η , for example, $l \approx \eta, 2\eta, 4\eta, \dots, L$) and the mixing times sum to a constant ($\sim \tau_L$) independent of κ by virtue of the Kolmogorov expression for τ_r . Thus a multi-scale flow acts as if it had an effective diffusivity, $\kappa_{\text{eff}} \approx D_{\text{eff}} \sim \Delta V L$. However, the latter interpretation is consistent with observation only for the averaged scalar dissipation. Large fluctuations of the dissipation rate reveal a breakdown of the 'effective diffusivity' view of mixing, and may be attributed to patchiness of the process where intermediate scales become transiently quiescent, leaving the mixing of large variations of the scalar entirely to the small-scale velocity. K41 'cascade' type theory does not tell us what these fluctuations are.

In turbulence theory, the discussions of intermittency and of the consequent deviations from K41 scaling often invoke the presence of "coherent structures"—for example, vortex filaments—amidst turbulent fluctuations^{9,10}. For the scalar, the excess (relative to K41) fluctuations on small scales and associated anomalous scaling exponents, as well as fluctuations of the local rate of mixing, can all be interpreted in terms of the structures revealed in experimental traces such as that shown in Fig. 3a. The saw-tooth appearance is indicative of a field organized into a transient array of plateaux, up to L in size, separated by sharp cliffs^{22,23}. (The scalar in Fig. 3a is constrained to have zero mean, which converts the plateaux to ramps). Remarkably, such events are sufficiently common (roughly one per integral scale and turnover time) that the normalized odd moments $S_{2n+1}(r)/S_2^{n+1/2}(r)$ which according to the K41 isotropization hypothesis should tend to zero as $(r/L)^{2/3}$ with decreasing r (or as $\text{Re}^{-1/2}$ for the corresponding derivative ratio) do so very slowly and possibly not at all^{7,23,25}.

Furthermore, numerical simulations^{16,17,31} have shown most decisively that the 'structures' in the scalar are not mere footprints of organized vorticity in shear flow, but are intrinsic to mixing. In the numerical simulation the velocity field is a structureless, artificial, gaussian process, yet the plateaux and cliff structures are clearly evident (Fig. 3b). The mechanism is intuitively appealing: turbulent mixing effectively expels whatever large-scale gradient there is, forming patches of nearly constant scalar field which are reconciled with the existing large-scale gradients by the formation of fronts (that is, intense gradient sheets).

Contrary to effective diffusivity notions, the sheets are sharp, and form transiently in the vicinity of the hyperbolic points of the large-scale flow (Fig. 7 of ref. 16) where 'jets' carrying high and low values of the scalar collide. Large and small scales are thereby directly coupled, producing large fluctuations on small scales and dominating high moments of the scalar structure function. If a static large-scale gradients is imposed, the fronts preferentially align normal to its direction and a violation of small-scale isotropization (for example, an anomalously large third order moment) is produced as well, so that odd moments computed numerically agree with experiment^{16,17}. The remarkable lesson here is that intermittency in scalar mixing is intrinsic, decoupled from that in the velocity field, and approachable via the simpler problem of advection by random flow pioneered by Batchelor¹² and Kraichnan^{14,15}—which we shall discuss next.

Kraichnan's model

Scalar transport and mixing is described by the following advection/diffusion equation:

$$\partial_t \Theta(\mathbf{r}, t) + \mathbf{v}(\mathbf{r}, t) \cdot \nabla \Theta(\mathbf{r}, t) = \kappa \nabla^2 \Theta(\mathbf{r}, t) \quad (1)$$

where Θ is the scalar field (that is, temperature or dye concentration), $\mathbf{v}(\mathbf{r}, t)$ is the velocity field, and κ is molecular diffusivity. In the absence of direct sources in the bulk of the fluid (which is often the case), the scalar is driven by external boundary conditions, which impose a time independent average $\langle \Theta(\mathbf{r}) \rangle$. We are interested in the deviations from this average, the fluctuations $\theta(\mathbf{r}, t) \equiv \Theta(\mathbf{r}, t) - \langle \Theta(\mathbf{r}) \rangle$. Rewriting equation (1) in terms of $\theta(\mathbf{r}, t)$ we find on the right-hand side an additional term $s(\mathbf{r}, t) \equiv -\mathbf{v}(\mathbf{r}, t) \cdot \nabla \langle \Theta(\mathbf{r}) \rangle$: the gradient forcing.

Alternatively, as a mathematical expedient, equation (1) can be considered with a gaussian random source. In reality, the velocity field is governed by the Navier–Stokes equation as appropriate to the experiment or natural phenomenon at hand. However, in the passive scalar model we replace real turbulent flow by an incompressible random velocity field, which Kraichnan^{14,15} took to be gaussian and δ -correlated in time. The velocity ensemble is then fully specified by the two-point correlator¹⁴:

$$\langle v_a(\mathbf{r}, t) v_b(\mathbf{0}, t') \rangle - \langle v_a(\mathbf{0}, t) v_b(\mathbf{0}, t') \rangle = D_{ab}^{(\gamma)}(\mathbf{r}) \delta(t - t') \quad (2)$$

where the magnitude of the symmetric tensor $D_{ab}^{(\gamma)}(\mathbf{r})$ varies as $\sim |\mathbf{r}|^{2-\gamma}$ and the dependence on the spatial indices ensures the incompressibility (that is, $\partial_r^a D_{ab}^{(\gamma)}(\mathbf{r}) = 0$). The averaging denoted by $\langle \dots \rangle$ is over the gaussian velocity ensemble. The two parameters of the model are the scaling index γ and (as the mathematical problem is well posed in space of any dimension greater than 1) the spatial dimension d , which will prove very useful analytically. To make contact with physical flows which have finite and scale-dependent correlation time $\tau(r)$, one compares $\int dt \langle (v(r, t) - v(0, 0))^2 \rangle$ with the (lagrangian) relative diffusivity of particles. Thus Richardson $r^{4/3}$ scaling of the relative diffusivity $D_{\text{eff}}(r)$ is recovered for $\gamma = 2/3$. The Batchelor, single-scale flow, limit is described by $\gamma \rightarrow 0$ and a 'rough' (non-differentiable) velocity field by $\gamma \rightarrow 2$.

Despite the relative simplicity of equations (1) and (2), the statistics of the scalar field is highly non-trivial. Kraichnan demonstrated in 1968 (ref. 19) that these equations lead to $\langle (\theta(\mathbf{r}, t) - \theta(\mathbf{0}, t))^2 \rangle \sim r^\gamma$, and thus, for $\gamma = 2/3$, consistency with KOC. Yet, he later surmised¹⁵ that all higher (even) moments of the scalar difference exhibit anomalous scaling and thus that the model equations (1, 2) also capture scalar intermittency.

The lagrangian view of mixing

The effect of advection (the $\mathbf{v} \cdot \nabla \theta$ term in equation (1)) can be dealt with mathematically by changing from the laboratory (or eulerian) coordinate system to lagrangian coordinates, which follow material elements in the flow. The diffusion term in these moving coordinates would appear intractable but, thanks to the linearity of the equation, the effect of diffusion may be correctly represented by including brownian motion in the lagrangian dynamics:

$$\frac{d}{dt} \theta(t) = s(\mathbf{p}(t), t) \quad (3)$$

$$\frac{d}{dt} \mathbf{p}(t) = \mathbf{v}(\mathbf{p}(t), t) + \boldsymbol{\eta}(t) \quad (4)$$

where $\theta(t)$ is the scalar concentration of the fluid element presently at position $\mathbf{p}(t)$ and $\boldsymbol{\eta}(t)$ is the Langevin noise $\langle \eta_a(t) \eta_b(t') \rangle = \kappa \delta_{ab} \delta(t - t')$ (introduced to represent the diffusion). In the absence of sources, θ is conserved along each trajectory, but the scalar field at the observation point $\mathbf{r}$ at time t is an average over

all the trajectories of the Langevin ensemble. The field at time t is related to the field at an earlier time t' by:

$$\theta(r, t) = \int dr' G(r, t | r', t') \theta(r', t') \quad (5)$$

with the Green's function representing the probability of trajectory (described by equation (4)) leaving from point r' at t' to arrive at r, t given by a path integral^{32–35}:

$$G(r, t | r', t') = \int_{\rho(t')=r'}^{\rho(t)=r} D\rho \exp \left[-\frac{1}{2\kappa} \int_{t'}^t dt'' (\dot{\rho} - v(\rho, t''))^2 \right] \quad (6)$$

The most likely trajectory arriving at the observation point (r, t) is the one which makes the 'action', in the exponent of equation (6), vanish. This zero-action path is exactly the deterministic lagrangian (L) trajectory governed by equation (4) without the Langevin noise. For short elapsed time $t - t'$, the action localizes these trajectories close to the L -trajectory so that equation (5) picks out the lagrangian pre-image of the observation point and equates the observed $\theta(r, t)$ to the θ of the pre-image. At longer times, the kernel G in equation (5) samples increasingly larger and larger volume of r' , making the observed $\theta(r, t)$ an average over the initial field and therefore decreasing its variance. This is how dissipation is represented by the path integral.

Having introduced the Green's function in equation (6), we construct a formal solution of equation (1) with sources which generalizes equation (5),

$$\theta(r, t) = \int_{-\infty}^t dt' \int dr' G(r, t | r', t') s(r', t').$$

The path integral lends itself to a semiclassical approximation: the transition probability expressed in equation (6) is dominated in the $\kappa \rightarrow 0$ limit by classical, minimal action trajectories (governed by the Euler–Lagrange equation) each contributing with the probability given by $\exp(-A_c/\kappa)$, A_c representing the action on the classical trajectory. For large $t - t'$, the classical trajectories diverge exponentially with the Lyapunov exponent (which measures the average rate of divergence) depending on the history of velocity gradient traversed by the L -trajectory. When the r' integral in equation (5) extends over a region much larger than the integral scale, the contribution of the sources self-averages, effectively cutting off the t' integral and defining a characteristic time for which the 'memory' of the sources persists. The latter quantity (on average) coincides with the mixing time t_* , but for each trajectory it depends on the velocity field in a well defined way enabling a quantitative analysis of the fluctuations.

The semiclassical approximation was used in ref. 35 to calculate the exponential tails in the single-point PDF $P(\theta)$ in the presence of a θ -gradient uniform over scales $\gg L$. Although the transport process between regions of size $\sim L$ is a random walk, large excursions in θ can occur for fluid parcels subject to little mixing. This is controlled by the trajectory-averaged Lyapunov exponent and, assuming gaussian statistics of $v(r, t)$, one can show that the probability distribution of mixing times is exponential³⁵. The exponential tails of $P(\theta)$ follow, in agreement with experiments^{24,26}. A δ -correlated velocity model is not a compromise with reality for processes that naturally occur on times much longer than the integral time τ_L , as is the case here. The tails of $P(\theta)$ can matter in practical problems. Consider the average rate of a thermally activated process (energy Δf) $\int d\theta \exp(-\Delta f/k_B \theta) P(\theta)$ or the probability that the pollution level exceeds some norm as it spreads from a source. A gaussian versus exponential distribution make very different predictions for the probability of rare events.

The statistical properties of $\nabla\theta$ may be computed in a similar fashion (in the Batchelor limit)^{35–37}. The key difference is that the gradient is amplified by the action of the strain along the path. This amplification is again controlled by the Lyapunov exponent along

the L -trajectory, and thus is in direct competition with dissipation acting through divergence of nearby paths expressed in equation (5). The PDF of the scalar dissipation $P(\nabla\theta^2)$ and the statistics of the Lyapunov exponents are discussed in refs 36 and 37 and can be compared with experimental data³⁸.

Multi-point correlators

Further information about the fluctuating scalar field may be obtained by studying multipoint correlators, $C_N = \langle \theta(r_1, t) \theta(r_2, t) \dots \theta(r_N, t) \rangle$ which involve simultaneous measurements of the field at N points. Unlike the traditional two-point structure function, $S_N(r)$, these objects contain not only the information about scaling properties but, through their dependence on the configuration of the measurement points, also probe the spatial structure of relevant fluctuations. Remarkably, these multipoint correlators also provide a convenient point of departure for the mathematical description of scalar statistics. An explicit path integral expression relating $C_N(t)$ to its value at earlier times follows readily from equations (5) and (6), but it is even more useful to consider evolution over very short times. For Kraichnan's δ -correlated model given in equation (2) the latter may be expressed in a differential form^{14,35}:

$$\partial_t \langle \theta(r_1, t) \dots \theta(r_N, t) \rangle = H_N \langle \theta(r_1, t) \dots \theta(r_N, t) \rangle + I_N(r_1, \dots, r_N, t) \quad (7)$$

with the so-called 'Hopf' evolution operator:

$$H_N \equiv - \sum_{i \neq j} [D_{ab}^{(j)}(r_i - r_j) + \kappa \delta_{ab} \partial_i^a \partial_j^b] \quad (8)$$

where the pair diffusivity $D_{ab}^{(j)}(r)$ defined in equation (2) appears upon performing the gaussian velocity average. The inhomogeneous source terms, $I_N(r_1, \dots, r_N, t)$, on the right-hand side depend only on the lower-order correlators ($N - 2$ and $N - 1$ for gradient forcing) so that equation (7) is a closed hierarchy which can be solved starting with $N = 2$. This equation was first written down and solved for $N = 2$ by Kraichnan¹⁴. Physically, the Hopf operator generalizes Richardson diffusion to many pairs of particles as they follow their lagrangian trajectories. Just as for ordinary diffusion, equation (7) is first order in time and second order in spatial derivatives. The same operator governs the evolution of a 'swarm' of N particles advected by random flow, and defines the transition probability from one configuration to another. The evolution operator H_N plays the role of the hamiltonian in the Schrodinger equation: this analogy with quantum mechanics is particularly evident in the path-integral representation.

In the presence of steady forcing, the correlators become stationary, the time derivatives vanish, and each C_N has to be determined in terms of the previous $C_{j < N}$ in the hierarchy by inverting the Hopf operator H_N . By analogy with the harmonic polynomials which solve $\nabla^2 \psi = 0$ and are necessary to satisfy the boundary conditions in electrostatics, homogeneous solutions or zero modes, $H_N \Psi = 0$, have to be added to each order to satisfy the boundary conditions, where the inertial range solution is matched to the large-scale flow. Thus the general solution has this form:

$$C_N = H_N^{-1} I_N + \sum_j a_j \Psi_{N,j} \quad (9)$$

Each term in equation (9) can be characterized in the inertial range by a scaling exponent λ , defined by $f(br_i) = b^\lambda f(r_i)$. The exponents for the inhomogeneous solutions $H_N^{-1} I_N$ are given by simple dimensional analysis, $\lambda_N^{(i)} = N\gamma/2$, and agree with KOC theory. However for the zero modes of H_N , the scaling exponents are non-trivial and have to be determined as nonlinear eigenvalues along with $\Psi_{N,j}$ themselves. (The index j further enumerates the behaviour of Ψ under symmetries such as reflections, rotations and permutations.) Each of the terms in equation (9) varies with scale as $(r/L)^{\lambda_{N,j}}$ and hence, in the inertial range $r \gg L$, the term with the smallest positive

exponent dominates, so that the structure function exponent is $\zeta_N = \min\{\lambda_N^{(d)}, \lambda_{N,j}\}$. Anomalous scaling or intermittency corresponds to $\zeta_N < \lambda_N^{(d)}$. Modes other than the one with smallest positive λ define the rate at which all the details of the integral scales decay away within the inertial range. (Negative $\lambda_{N,j}$ also exist, and describe how the influence of dissipation or local source decays as r increases.) This “zero mode” mechanism of anomalous scaling was recognized and investigated independently in refs 18–20.

In identifying ζ_N with the dominant scaling exponent of C_N , we have tacitly used the fact that the moments of the structure function, S_N , can be obtained simply by ‘fusing’ points together. This can be demonstrated explicitly. Furthermore, whenever a pair of points get very close, say $\rho = r_{12}$ much less than all other r_{ij} , one finds from the local analysis using equation (7) that the ρ dependence of C_N follows $S_2(\rho)$ (refs 37, 39, 40). This allows the scaling of C_N to be related to the scaling of the correlators of dissipation at different points³⁷ providing a link between the Hopf equation approach and the traditional parametrization (since Kolmogorov’s 1962 theory^{10,11}) of intermittency¹⁰ in terms of the fluctuations in the dissipation rate. However, it is not possible to circumvent the need for global solution of the $(N-1)$ d dimensional Hopf equation for C_N in computing the exponents by working with the structure function (depending on the single distance r). There is no closed equation with any of the $r_{ij} = 0$, as is evident in the path integral formulation: bringing the r_i observation points together, for example, $\langle \theta^N(r_1, t) \rangle$, does not eliminate the need to integrate over N distinct trajectories (now ending at the same r_1).

The Hopf equation ignores the temporal correlations of the velocity field, $\tau_r \approx r^{2/3}$ except in so far as they enter $D_{ab}(r)$. However at high Reynolds number fluid mechanics requires on average that $\tau_r \Delta_r V \approx r$, that is, the displacement of two points in one correlation time is of order their spacing. Thus, if one takes the $\tau \rightarrow 0$ limit while maintaining $\tau_r \Delta_r V \approx r$ scaling, the evolution equation (7) must be scale invariant and hence the scaling dimension of the H_N should be 0. Because temporal correlations are defined in a lagrangian frame, it is not possible to provide a simple gaussian model in eulerian coordinates analogous to equation (2). One approach is to eschew a model velocity field completely and model the Hopf operator directly, using symmetries and known limits as a guide^{20,40}. A similar methodology has proved effective for the strong interactions in high-energy physics. Consequences of this approach are contained in the following section, but the discussion so far should already convince the reader that the scaling exponents for C_N do depend on the precise form of H_N , and thus in principle on more (or other) parameters than γ and d which were defined in equation (2).

Anomalous scaling

For general parameter values, analytic solution of equation (7) is hopeless; but by moving away from the physical regime ($d = 3$, $\gamma = 2/3$), perturbative expansions for exponents can be obtained, which definitively establish the intermittency corrections to KOC. Three limits have been explored: (1) the diffusion limit $2 - \gamma \ll 1$ (corresponding to a rough velocity field) studied by Gawedsky and Kupiainen¹⁸; (2) the large spatial dimension limit $d \gg 1$ of Chertkov *et al.*¹⁹ and (3) the near-Batchelor limit $\gamma \ll 1$ (corresponding to a smooth velocity field) studied by Pumir and the authors^{20,40,41}. In the former two limits the expansion is about a gaussian solution where the $2n$ point correlators factorize pairwise. The two limits can be combined into a single expression, $\zeta_{2n} - \zeta_{2n}^K = -2(2 - \gamma)n(n-1)/(d+2) + o(2 - \gamma)^2/d^2$. The nearly smooth field case, $\gamma \ll 1$, relies on the integrability of the Batchelor limit^{20,40,42} resulting from the underlying high symmetry of H_N in that case. However, the dependence on small γ is non-analytic, and the required singular perturbation theory has only been performed for $N = 3$. The third moment, which appears in the presence of large-scale anisotropy, is the lowest correlator with non-trivial scaling: $\zeta_3 = 1 + o(\gamma^{3/2}) < \zeta_3^K = 1 + \gamma$. In the $2 - \gamma \ll 1$ limit⁴³, $\zeta_3 = \zeta_3^K - 2\gamma/5$ is also anomalous. Furthermore, Pumir^{43,44} obtained a numerical solution of the Hopf equation for $N = 3$ (and all γ in $d = 2, 3$) yielding in particular $\zeta_3 = 1.38$ for the physically relevant case of $\gamma = 2/3$ and $d = 3$.

Frisch, Mazzino, Noullez and Vergassola⁴⁵ implemented the lagrangian path integral representation of equation (4) as a numerical Monte Carlo algorithm for solving equations (1) and (2), and calculated ζ_{2n} for C_{2n} by following $2n - 1$ points (the centre of mass is irrelevant) backwards in time. Results for $n = 2$ appear in Fig. 4. In addition, for a velocity field with realistic temporal correlations and $\gamma = 2/3$, it has been established⁴⁶ in two dimensions that the ζ_{2n} approach a constant value of 1.4 with increasing n . The saturation of the exponent would be expected if intermittency was due to front-like structures, and was predicted by the semi-classical analysis^{47,48}.

Returning to the third order moment, we note that it is interesting both as a lowest-order anomalous correlator and because its anomaly carries a clear physical significance: it indicates that the anisotropy which is always present on large scales finds its way to the small scales. Enforcing isotropy on the large scales does not eliminate this phenomenon, but merely requires a multi-scale correlator to make it visible again. Experiments in many geometries^{7,8} quite generally find $\zeta_3 \approx 1$ —a much stronger anomaly than the 1.38 found in the δ -correlated model. The disagreement is not altogether surprising given that, as we have

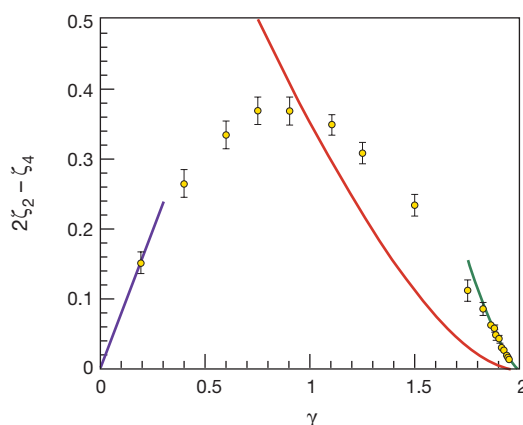


Figure 4 Anomalous scaling exponent, $2\zeta_2 - \zeta_4$, versus scaling index γ (see equation (2)). The structure function, was obtained from the lagrangian Monte Carlo simulation⁴⁵ of

the Kraichnan model (circles) and is compared with analytic perturbation theory^{18,41} of the Hopf equation (blue, green) and the prediction of the linear Ansatz¹⁵ (red).

already remarked, the latter model does not properly capture the lagrangian correlation and that the scaling exponents are non-universal eigenvalues which depend on the details of the evolution operator H_N . The class of phenomenological evolution operators introduced in ref 20 allows for stronger intermittency. The theory also produces a prediction for the full configuration dependence of $C_3(r_1, r_2, r_3)$ which has been compared to the experimental measurements in a wind tunnel²⁵ (with an imposed temperature gradient). The configuration dependence of C_3 is the simplest measure of the 'geometry' of the intermittent structures. Figure 5 illustrates how the contribution (to C_3) of randomly distributed, but preferentially oriented, fronts depends on the geometry of the triangle (within a naive model where the fronts are sharp, straight and of unitary height). To the extent that $\zeta_3 \approx 1$ (expected if the fronts are sharp and unitary), and the configuration dependence based on the 'naïve' model qualitatively conforms with the observation, the front interpretation stands. Yet both the explicit calculation C_3 and the observation²⁵ exhibit quantitative deviations from the simple front model, indicating that the height of the fronts, their width and spatial extent all have non-trivial distribution⁴⁶.

Higher multipoint correlators of θ and $\partial\theta$ can be used to study the latter in detail. It would be particularly interesting to examine the four-point correlator $\langle\partial\theta(r_1)\partial\theta(r_2)\partial\theta(r_3)\partial\theta(r_4)\rangle$. If the fluctuations are dominated by sheets, this four-point object should be maximal when the points are coplanar⁴⁹. The decay of the correlator as one of the points leaves the plane would quantify the width of the sheets which may be expected to depend on the scale, controlled by the separation of the other three points.

Lessons of scalar turbulence

The study of nonequilibrium problems has lagged behind systems in thermal equilibrium because there is no way, analogous to averaging over the Boltzmann distribution, to compute equal time averages. Equilibrium statistical mechanics would be a very different subject if it were necessary to time-average solutions to Newton's equations. Thus, the recent progress in understanding passive scalar turbulence was grounded on the computationally useful representation of the stationary statistics. The lagrangian path integral formulation mapped the problem into statistical mechanical form. The Hopf equation (7) made possible systematic perturbative calculations of the multipoint correlators. Qualitatively new is the mechanism for generating anomalous scaling via zero modes of the Hopf operator which associates intermittency effects with the existence of statistically stationary unforced solutions in the inertial range. The operator is inherently multidimensional, and the statistics of the multipoint configurations are essential to understand intermittency. The theory naturally allows for an infinity of subdominant exponents that parametrize how all manner of large-scale anisotropies decay away in the inertial range. There is good evidence that

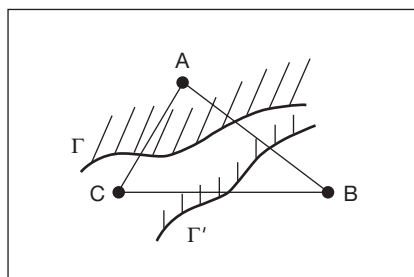


Figure 5 Contribution of fronts and configuration dependence of the three-point correlator. If Γ and Γ' are fronts separating $\theta = 1$ from $\theta = -1$, the former contributes +1 to the C_3 correlator, while the latter contributes -1. The resulting value of C_3 depends on the shape of the ABC triangle.

the anomalous scaling exponents reach saturation for large N as would be the case with sharp unitary fronts dominating large fluctuations as in Burger's "turbulence"¹⁰.

Exponents appear as nonlinear eigenvalues, and thus depend on all details of the velocity statistics embodied in the Hopf operator. This dependence on the velocity ensemble makes the exponents for passive scalar turbulence non-universal, and unless the statistics of the turbulent velocity itself proves to be fully universal (that is, independent of the forcing and boundary conditions) in an experimentally attainable regime, this lack of universality should manifest itself in observations.

The lagrangian path integral representation has provided both a useful analytic tool and a new numerical simulation approach. It has been used to produce considerable analytic insight into dissipation statistics³⁷ and into the nature of the direct and inverse cascade⁵⁰, as well as a reliable numerical calculation of the structure function exponent^{45,46}. Also significant is the emergence of the $1/d$ expansion¹⁹ in the context of turbulence⁵¹. The most immediate application for these tools is in the context of the magnetic dynamo problem^{34,49}, where the amplification of the magnetic field vector is in a certain way analogous to the behaviour of the passive scalar gradient.

It is useful to contrast passive scalar turbulence with critical phenomena at second-order phase transitions⁵², as both problems exhibit scale-invariant correlations. The obvious distinction is that the scaling behaviour in turbulence is confined to inertial scales ($r < L$) and the correlations grow with r/L rather than decrease at long distance as in critical phenomena. The universality in critical theories is embodied by the fixed-point equation for the effective free energy at long wavelength, and that only two exponents suffice to determine all the others. Scaling exponents are additive as long as the fields entering the correlator are not at the same point and develop 'anomalies' when the observation points are fused. Curiously, the situation is opposite for the passive scalar, where addition of an observation point to a correlator picks up a new, higher-order exponent, whereas points fuse with impunity.

The lessons and methods of the passive scalar problem have implications for hydrodynamic turbulence, even though the defining equations are no longer linear. Subgrid modelling or Reynolds stress closures which seek to parametrize the unresolved small scales have a long history in turbulence modelling¹¹, and multipoint correlators might point to better representation of the interactions between scales. Furthermore, the tendency towards the formation of structures for the passive scalar may be robust enough to carry over to the velocity field. In particular, scalar gradient sheets and their alignment with the large-scale gradient may have an analogue in the transient appearance of vortex sheet structures in shear flow^{53,54}. The latter would result in the persistence of anisotropy on small scales which becomes only stronger for higher-order moments.

Ultimately, the most significant lesson of the passive scalar problem is the fundamental importance of the multipoint correlators. The configuration dependence of C_N is both measurable and dynamically significant. Prior work on turbulence concentrated on either two-point statistics and its scaling or on 'snapshots' of the velocity or scalar fields and identification of spatial 'structures'. Multipoint correlators naturally fill the gap between the two approaches; they provide new, physically interesting observables for the experiments and direct numerical simulations, and ultimately new prospects for the quantitative understanding of turbulence. □

1. Pasquill, F. & Smith, F. B. *Atmospheric Diffusion* (Ellis Horwood, Chichester, 1983).
2. Dibble, R. W., Warnatz, J. & Maas, U. *Combustion: Physical and Chemical Fundamentals, Modelling and Simulations, Experiments, Pollutant Formation* (Springer, New York, 1996).
3. Proctor, M. R. E. & Gilbert, A. D. (eds) *Lectures on Solar and Planetary Dynamos* 117 (Cambridge Univ. Press, Cambridge, 1994).
4. Berg, H. C. Bacterial microprocessing. *Cold Spring Harbor Symp. Quant. Biol.* **55**, 539–545 (1990).
5. Atema, J. Chemical signals in the marine environment: dispersal, detection, and temporal signal analysis. *Proc. Natl Acad. Sci.* **92**, 62–66 (1995).

6. Mafrano, A. & Carde, R. T. Fine-scale structure of pheromone plumes modulates upwind orientation of flying moths. *Nature* **369**, 142–144 (1994).
7. Sreenivasan, K. R. On local isotropy of passive scalars in turbulent shear flows. *Proc. R. Soc. Lond. A* **434**, 165–182 (1991).
8. Warhaft, Z. Passive scalars in turbulent flows. *Annu. Rev. Fluid Mech.* **32**, 203–240 (2000).
9. Sreenivasan, K. & Antonia, R. A. The phenomenology of small scale turbulence. *Annu. Rev. Fluid Mech.* **29**, 435–472 (1997).
10. Frisch, U. *Turbulence: The legacy of AN Kolmogorov* (Cambridge Univ. Press, Cambridge, 1995).
11. Monin, A. S. & Yaglom, A. M. *Statistical Fluid Mechanics* (MIT Press, Cambridge, 1975).
12. Tennekes, H. & Lumley, J. A. *First Course in Turbulence* (MIT Press, Cambridge, 1972).
13. Edouard, S. *et al.* The effect of small-scale inhomogeneities on ozone depletion in the Arctic. *Nature* **384**, 444–447 (1996).
14. Kraichnan, R. H. Small-scale structure of a scalar field convected by turbulence. *Phys. Fluids* **11**, 945 (1968); Convection of a passive scalar by a quasi-uniform random straining field. *J. Fluid Mech.* **64**, 737–762 (1974).
15. Kraichnan, R. H. Anomalous scaling of a randomly advected passive scalar. *Phys. Rev. Lett.* **72**, 1016–1019 (1994).
16. Holzer, M. & Siggia, E. D. Turbulent mixing of a passive scalar. *Phys. Fluids* **6**, 1820–1837 (1994).
17. Pumir, A. A numerical study of the mixing of a passive scalar in 3 dimensions in the presence of a mean gradient. *Phys. Fluids* **6**, 2118–2132 (1994).
18. Gawędzki, K. & Kupiainen, A. Anomalous scaling of the passive scalar. *Phys. Rev. Lett.* **75**, 3834–3837 (1995).
19. Chertkov, M., Falkovich, G., Kolokolov, I. & Lebedev, V. Normal and anomalous scaling of the fourth-order correlation function of randomly advected passive scalar. *Phys. Rev. E* **52**, 4929–4941 (1995).
20. Shraiman, B. I. & Siggia, E. D. Anomalous scaling of a passive scalar in a turbulent flow. *C.R. Acad. Sci.* **321**, 279–284 (1995).
21. Antonia, R. A. & van Atta, C. W. Structure functions of temperature fluctuations in turbulent shear flows. *J. Fluid Mech.* **84**, 561–580 (1978).
22. Gibson, C., Freije, C. & McConnell, S. Structure of sheared turbulent fields. *Phys. Fluids* **20**, S156–S167 (1977).
23. Mestayer, P. G. Local isotropy and anisotropy in a high-Reynolds-number turbulent boundary layer. *J. Fluid Mech.* **125**, 475–503 (1982).
24. Jayesh & Warhaft, Z. Probability distributions of a passive scalar in grid generated turbulence. *Phys. Rev. Lett.* **67**, 3503–3506 (1991).
25. Mydlarski, L., Pumir, A., Shraiman, B. I., Siggia, E. D. & Warhaft, Z. Structure and multi-point correlators for turbulent advection. *Phys. Rev. Lett.* **81**, 4373–4376 (1998).
26. Gollub, J. P., Clarke, J., Gharib, M., Lane, B. & Mesquita, O. N. Fluctuations and transport in a stirred fluid with a mean gradient. *Phys. Rev. Lett.* **67**, 3507–3511 (1991).
27. Ottino, J. M. *The Kinematics of Mixing: Stretching, Chaos, and Transport* (Cambridge Univ. Press, New York, 1989).
28. Ott, E. *Chaos in Dynamical Systems* (Cambridge Univ Press, New York, 1994).
29. Williams, B. S., Marteau, D. & Gollub, J. P. Mixing of a passive scalar in magnetically forced two-dimensional turbulence. *Phys. Fluids* **9**, 2061–2080 (1997).
30. Miller, P. L. & Dimotakis, P. E. Measurements of scalar power spectra in high Schmidt number turbulent jets. *J. Fluid Mech.* **308**, 129–146 (1996).
31. Chen, S. & Kraichnan, R. H. Simulations of a randomly advected passive scalar field. *Phys. Fluids* **68**, 2867–2884 (1998).
32. Feynman, R. & Hibbs, A. *Quantum Mechanics and Path Integrals* (McGraw Hill, New York, 1965).
33. Drummond, I. T. Path-integral for turbulent diffusion. *J. Fluid Mech.* **123**, 59–68 (1982).
34. Molchanov, S., Ruzmaikin, A. & Sokolov, A. Kinematic dynamo. *Sov. Phys. Usp.* **28**, 307–327 (1985).
35. Shraiman, B. I. & Siggia, E. D. Lagrangian path integrals and fluctuations in random flow. *Phys. Rev. E* **49**, 2912–2927 (1994).
36. Bernard, D., Gawędzki, K. & Kupiainen, A. Slow modes in passive advection. *J. Stat. Phys.* **90**, 519–569 (1997).
37. Chertkov, M., Falkovich, G. & Kolokolov, I. Intermittent dissipation in PS turbulence. *Phys. Rev. Lett.* **80**, 2121–2124 (1998).
38. Buck, K. A. & Dahm, W. J. A. Experimental study of the fine scale structure of conserved scalar mixing in turbulent shear flows. Part I. *J. Fluid Mech.* **317**, 21–71 (1996).
39. Fairhall, A., Gat, O., L'vov, V. & Procaccia, I. Anomalous scaling in a model of passive scalar advection. *Phys. Rev. E* **53**, 3518–3535 (1996).
40. Shraiman, B. I. & Siggia, E. D. Symmetry and scaling of turbulent mixing. *Phys. Lett.* **77**, 2463–2466 (1996); Anomalous scaling of passive scalar near Batchelor limit. *Phys. Rev. E* **57**, 2965–2977 (1998).
41. Pumir, A., Shraiman, B. I. & Siggia, E. D. Perturbation theory for the δ -correlated model of passive scalar advection near the Batchelor limit. *Phys. Rev. E* **55**, R1263–R1266 (1997).
42. Balkovsky, E., Chertkov, M., Kolokolov, I. & Lebedev, V. 4-th order correlation function of a randomly advected passive scalar. *JETP Lett.* **61**, 1049–1054 (1995).
43. Pumir, A. Anomalous scaling behaviour of a passive scalar in the presence of a mean gradient. *Europhys. Lett.* **34**, 25–29 (1996).
44. Pumir, A. Structure of the three-point correlation function of a passive scalar in the presence of a mean gradient. *Phys. Rev. E* **57**, 2914–2929 (1998).
45. Frisch, U., Mazzino, A., Noullez, A. & Vergassola, M. Lagrangian method for multiple correlations in passive scalar advection. *Phys. Fluids* **11**, 2178–2186 (1999).
46. Celani, A., Lanotte, A., Mazzino, A. & Vergassola, M. Universality and saturation of intermittency in passive scalar turbulence. *Phys. Rev. Lett.* **84**, 2385–2388 (2000).
47. Chertkov, M. Instanton for random advection. *Phys. Rev. E* **55**, 2722–2735 (1997).
48. Balkovsky, E. & Lebedev, V. Instanton for the Kraichnan random advection model. *Phys. Rev. E* **58**, 5776–5795 (1998).
49. Chertkov, M., Falkovich, G., Kolokolov, V. & Vergassola, M. Small scale magnetic dynamo. *Phys. Rev. Lett.* **83**, 4065–4068 (1999).
50. Chertkov, M., Kolokolov, L., Vergassola, M. Inverse versus direct cascades in turbulent advection. *Phys. Rev. Lett.* **80**, 512–515 (1998).
51. Fournier, J.-D., Frisch, U. & Rose, H. Infinite-dimensional turbulence. *J. Phys. A* **11**, 187–198 (1978).
52. Wilson, K. G. The renormalization group and critical phenomena. *Rev. Mod. Phys.* **55**, 583–600 (1983).
53. Pumir, A. & Shraiman, B. I. Persistent small scale anisotropy in homogeneous shear flows. *Phys. Rev. Lett.* **75**, 3114–3117 (1995).
54. Warhaft, Z. & Garg, S. On the small scale structure of simple shear flow. *Phys. Fluids* **10**, 662–673 (1998).

Acknowledgements

The authors have greatly benefited from interactions with colleagues, and are particularly grateful to S. Chen, K. Sreenivasan, P. Tabeling, Z. Warhaft and L. Mydlarski for contributing to the figures.

Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution

Chikashi Toyoshima^{*†}, Masayoshi Nakasako^{*†‡}, Hiromi Nomura^{*} & Haruo Ogawa^{*}

^{*} Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan

[†] The Harima Institute, The Institute of Physical and Chemical Research, Sayo-gun, Hyogo 679-5143, Japan

[‡] PRESTO, Japan Science and Technology Corporation, Kawaguchi 332-0012, Japan

Calcium ATPase is a member of the P-type ATPases that transport ions across the membrane against a concentration gradient. Here we have solved the crystal structure of the calcium ATPase of skeletal muscle sarcoplasmic reticulum (SERCA1a) at 2.6 Å resolution with two calcium ions bound in the transmembrane domain, which comprises ten α -helices. The two calcium ions are located side by side and are surrounded by four transmembrane helices, two of which are unwound for efficient coordination geometry. The cytoplasmic region consists of three well separated domains, with the phosphorylation site in the central catalytic domain and the adenosine-binding site on another domain. The phosphorylation domain has the same fold as haloacid dehalogenase. Comparison with a low-resolution electron density map of the enzyme in the absence of calcium and with biochemical data suggests that large domain movements take place during active transport.

The calcium pump of sarcoplasmic reticulum was first identified in the 'relaxing factor' of muscle contraction and gave rise to the calcium theory¹ that Ca^{2+} is a fundamental and ubiquitous factor in the regulation of intracellular processes. It is an integral membrane protein of relative molecular mass 110,000 (M_r 110K) and pumps Ca^{2+} released in muscle cells during contraction back into the sarcoplasmic reticulum using the chemical energy of ATP, thereby relaxing muscle cells. This pump is therefore an ATPase and a representative member of a group of enzymes called P-type ATPases; the name came from the fact that they are autophosphorylated (at Asp 351 with Ca^{2+} -ATPase) during the reaction cycle (for recent reviews, see refs 2–4). P-type ATPases are all ion pumps of crucial importance, for example Na^+K^+ -ATPase and gastric H^+K^+ -ATPase.

The amino-acid sequences of P-type ATPases have been known for some time, and secondary-structure prediction has proposed models with ten transmembrane helices (M1–M10)⁵. Extensive mutational studies have been carried out to identify the amino-acid residues that are critical for ion transport (for example, see ref. 3). Negatively charged residues on four putative transmembrane helices (M4–M6 and M8) have been postulated to form high-affinity binding sites⁶. The evolutionary connections with other enzyme families have been a long-standing issue, because P-type ATPases lack the P-loop commonly found in other ATPases and GTPases⁷. Nevertheless, marked homology with L-2-haloacid dehalogenase has been pointed out for the catalytic core of the cytoplasmic domain⁸.

Direct structural information has so far been limited because the crystals that had been obtained were useful only for electron microscopy. Two types of crystal have been obtained for sarcoplasmic reticulum Ca^{2+} -ATPase: tubular crystals formed in the absence of Ca^{2+} and the presence of decavanadate⁹, and three-dimensional microcrystals formed in the presence of millimolar Ca^{2+} (refs 10, 11). From electron microscopy of these crystals, it is clear that the enzyme has a large cytoplasmic 'headpiece' connected to the transmembrane region by a short stalk segment^{12,13}, and that the headpiece appears to split on Ca^{2+} binding to the membrane domain¹⁴.

Here we have improved the crystallization conditions and succeeded in growing crystals suitable for X-ray crystallography. We describe the structure of the sarcoplasmic reticulum calcium pump

with two Ca^{2+} ions in the transmembrane binding sites at 2.6 Å resolution. The atomic model explains the 8 Å resolution map obtained from tubular crystals¹³, suggesting that large concerted domain motions take place during active transport.

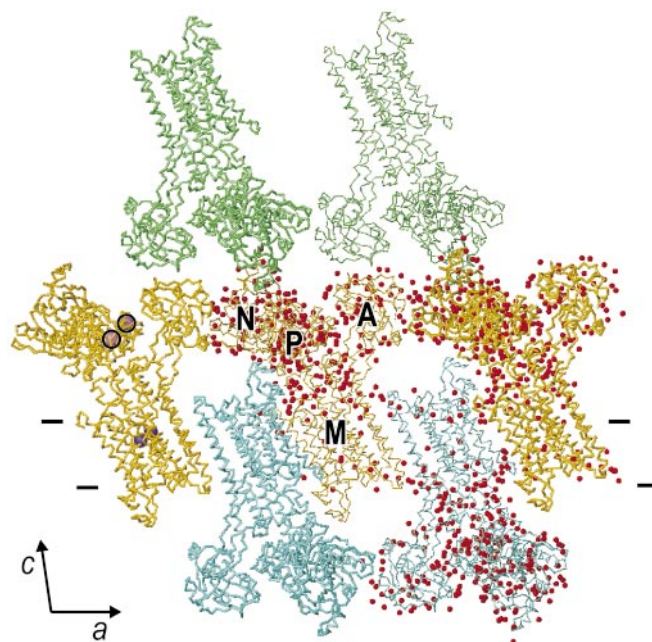


Figure 1 Crystal packing of Ca^{2+} -ATPase in the plate-like crystals. C α traces with the a axis horizontal and the b axis normal to the plane of the paper. Thin lines represent the molecules offset by a half unit cell along the b axis owing to the crystal symmetry. Because the crystal is made of stacks of membranes, the lipid bilayers extend parallel to the ab plane (normal to the plane of the paper) but may well be distorted around the protein molecules. Horizontal bars show rough estimates of the positions of the membrane surface from the distribution of hydration water. Three cytoplasmic domains (A, N and P) are identified as in Fig. 2. Locations of two Ca^{2+} ions are indicated by spheres in violet in the transmembrane domain (M). Those of lanthanides (La^{3+} and Tb^{3+}) in derivative crystals (see Table 1) are marked by circles. No binding of lanthanides was detected in the transmembrane region. The distance between adjacent layers measures 145.7 Å.

Table 1 Summary of crystallographic analysis

Data set*	Concentration (mM)	Resolution (Å)	Completeness (%)	$R_{\text{merge}}^{\dagger}$ (%)	Redundancy	I/σ	$R_{\text{iso}}^{\ddagger}$ (%)	No. of sites	Phasing power§
Native (eight crystals)		2.60	99.9	6.6 (25.9)#	27.2	29.3 (4.3)#			
PIP	1	3.20	98.0	6.1 (17.8)	3.4	23.7 (7.7)	10.2	1	0.40
$\text{K}_2\text{Pt}(\text{NO}_2)_4$	2	2.80	99.5	6.0 (21.5)	3.3	24.7 (6.7)	12.6	4	1.01
	0.4	3.10	99.8	6.9 (20.1)	3.8	25.6 (5.4)	8.7	4	1.02
$\text{Pt}(\text{NH}_3)_4(\text{NO}_3)_2$	1	3.10	99.8	6.8 (20.7)	3.6	16.8 (5.0)	5.9	2	0.45
$\text{K}_2\text{Pt}(\text{CN})_6$	2	3.20	94.5	4.0 (13.3)	3.4	11.3 (5.6)	11.0	3	0.34
$\text{Tb}(\text{NO}_3)_3$	2	3.15	98.2	5.6 (18.1)	3.4	10.3 (4.2)	11.5	2	1.00
(soaked 90 h)	1	3.00	99.9	6.0 (18.2)	2.7	9.2 (4.7)	8.0	2	0.85
$\text{La}(\text{NO}_3)_3$	2	3.20	98.0	4.0 (13.0)	3.5	15.1 (6.0)	6.9	2	0.82
TNP-AMP	0.5	4.00	94.7	10.8 (18.1)	2.8	6.0 (4.1)	9.9		
Refinement statistics									
Resolution range	No. of reflections	No. of protein atoms	No. of water molecules	No. of Ca^{2+} atoms	$R_{\text{crist}}^{\parallel}$ (%)	$R_{\text{free}}^{\P}$ (%)	R.m.s. bond lengths (Å)	R.m.s. bond angle (°)	
15–2.6 Å	48,373 (92.5%)	7,673	276	2	25.0	30.7	0.008	1.4	

* Diffraction data were collected at SPring-8 (beamline BL41XU ($\lambda = 0.800$ Å) and BL44B2 ($\lambda = 0.890$ Å and 1.000 Å) for native crystals and BL44B2 ($\lambda = 0.890$ Å) for derivatives).
† $R_{\text{merge}} = \sum_i \sum_h |I_i(hkl) - \langle I(hkl) \rangle| / \sum_i \sum_h I_i(hkl)$.
‡ $R_{\text{iso}} = \sum_h |F_{\text{obs}}(hkl) - \langle F_{\text{obs}}(hkl) \rangle| / \sum_h |F_{\text{obs}}(hkl)|$.
§ Defined as the ratio of the r.m.s. value of the heavy atom structure factor amplitudes and the r.m.s. value of the lack-of-closure error.
|| $R_{\text{crist}} = \sum_h |F_{\text{obs}}(hkl) - F_{\text{calc}}(hkl)| / \sum_h |F_{\text{obs}}(hkl)|$.
¶ Same as R_{crist} , but calculated with 10% of data set aside for refinement.
The numbers in the parentheses refer to those in the last resolution shell (2.68–2.60 Å for the native data set).

Figure 2 Architecture of the sarcoplasmic reticulum Ca^{2+} -ATPase. α -Helices are represented by cylinders and β -strands by arrows, as recognized by DSSP⁴⁶. Cylinders are not used for one-turn helices. Colour changes gradually from the N terminus (blue) to the C terminus (red). Three cytoplasmic domains are labelled (A, N and P). Transmembrane helices (M1–M10) and those in domains A and P are numbered. The model is orientated so that transmembrane helix M5 is parallel to the plane of the paper. The model in the right panel is rotated by 50° around M5. The M5 helix is 60 Å long and

serves as a scale. Several key residues are shown in ball-and-stick, and TNP-AMP by CPK. D351 is the residue of phosphorylation. Two purple spheres represent Ca^{2+} in the transmembrane binding sites. The binding sites for phospholamban (PLN)¹⁶ and thapsigargin (TG)¹⁷ are marked, as are major digestion sites for trypsin⁵ (T1 and T2) and proteinase K²⁸ (PrTK). The arrow specifies the direction of view in Fig. 6b. Figure prepared with MOLSCRIPT⁴⁷.

Structure determination

The crystals were grown at pH 6.1 in the presence of 10 mM CaCl_2 and exogenous lipid. Electron microscopy of these crystals showed that the protein molecules were indeed embedded in lipid bilayers (data not shown). They were very thin (typically $< 20 \mu\text{m}$); but each of them allowed a full data set to be collected at SPring-8. The structure was solved by standard multiple isomorphous replacement (MIRAS) methods and refined to 2.6 Å resolution. The quality of the diffraction data and the refinement statistics are given in Table 1. The atomic model included about 250 water molecules in the cytoplasmic and the luminal regions and about 30 in the transmembrane region. The distribution of hydration water molecules indicated approximately the position of the membrane boundary, which was roughly parallel to the a axis (Fig. 1).

Architecture of the sarcoplasmic reticulum Ca^{2+} -ATPase

The overall architecture of the sarcoplasmic reticulum Ca^{2+} -ATPase with bound Ca^{2+} is shown in Fig. 2. All 994 amino acid residues¹⁵ were identified in the electron-density maps (see Fig. 3a, b for the initial maps). The molecule fits in a box of $100 \text{ Å} \times 80 \text{ Å} \times 140 \text{ Å}$. The cytoplasmic headpiece has a split appearance consisting of three well separated domains (designated as A, N and P). The phosphorylation

site, Asp 351, is located in domain P, and, as described below, the adenosine moiety of nucleotide (here we used 2',3'-O-(2,4,6-trinitrophenyl)-AMP (TNP-AMP)) is bound to domain N.

The transmembrane region (M) comprises ten α -helices (M1–M10), the arrangement of which is shown in Fig. 3a. There is a clear segregation between M1–M6 and M7–M10, consistent with the lack of M7–M10 helices in bacterial type I P-type ATPases (for example, ref. 2). The lengths of helices and the inclination to the membrane vary substantially (Fig. 2). Some of them (M2 and M5) are very long ($\sim 60 \text{ Å}$) and fairly straight, but some are unwound (M4 and M6) or kinked (M10) at the middle of the membrane. The helices that have been presumed to form Ca^{2+} -binding sites (M4–6, M8) show very different folding patterns, although similarity among them was expected from mutation studies³. M5 is located in the centre of the molecule and extends from the luminal surface of the membrane to the centre of the cytosolic domain P, looking like the 'centre mast' of the enzyme. M2 and M3 are long helices that are rather isolated. The implication of this feature is discussed below. These long helices are continuous across the membrane surface and therefore treated as single helices, including the region previously assigned to 'stalk helices'⁵. Luminal loops are short except for the one connecting M7 and M8 (L78, ~ 35 residues). In

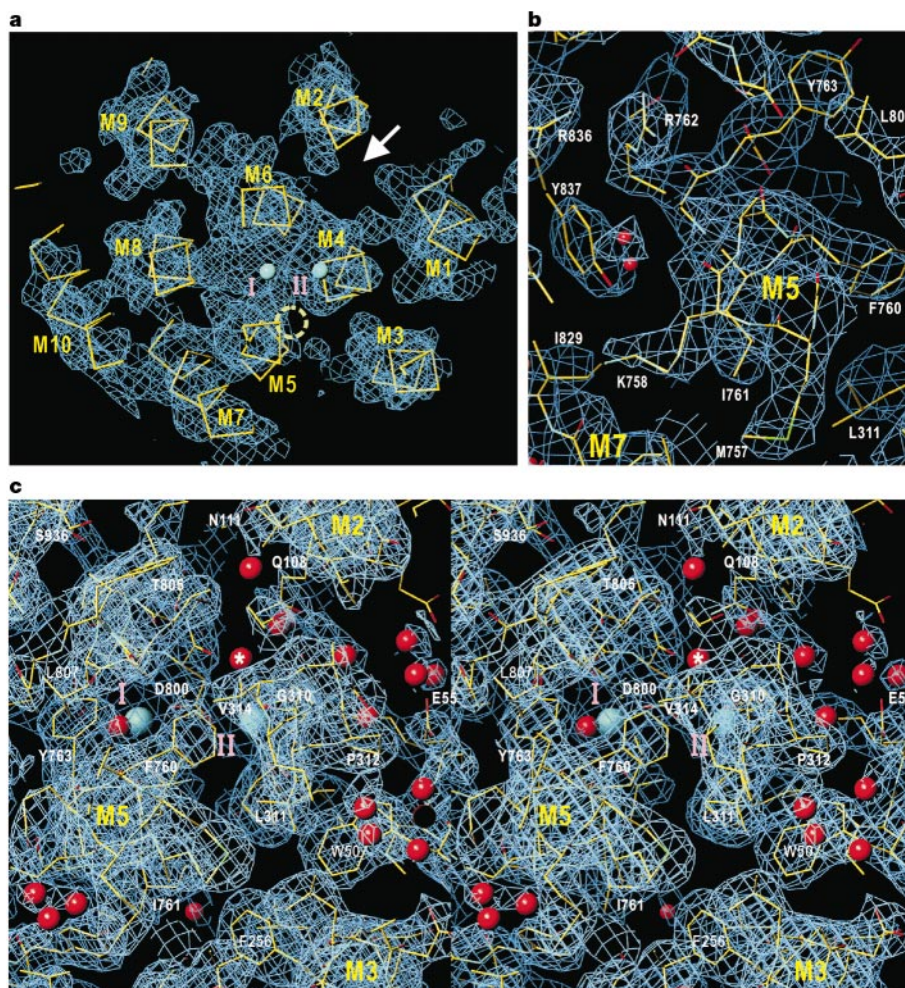


Figure 3 The arrangement of transmembrane helices (**a**), an enlarged view of a part of M5 helix (**b**) and water accessible cavities on the cytoplasmic surface (in stereo, **c**). The viewing direction is roughly normal to the membrane from the cytoplasmic side with domain A on the right hand side (**a,c**). The luminal half including the Ca^{2+} -binding sites is shown in **a** and the cytoplasmic half in **c**. Cyan spheres represent two Ca^{2+} ions (I and II) and red spheres represent water molecules. A solvent flattened map is used in **a** and **b**

(calculated at 2.8 Å resolution and contoured at 1.2σ), and a $2|F_o| - |F_c|$ composite omit map prepared with CNS⁴⁵ (calculated at 2.6 Å resolution and contoured at 1.5σ) in **c**. The arrow in **a** indicates the viewing direction in Fig. 4b. Circle in **a** indicates a likely position of the outlet of Ca^{2+} pathway. The water molecule found nearest to the Ca^{2+} -binding sites is marked (asterisk in **c**). Note that only M2, M3 and M5 are labelled in **c**.

Fig. 2, the major sites for protease digestion and the binding sites of phospholamban¹⁶ and thapsigargin¹⁷ are also identified.

Structure of the Ca^{2+} -binding sites

The M4–M6 and M8 helices surround two high-density peaks (Fig. 4), which we identify as two bound Ca^{2+} ions and not water. Our reasons are fourfold. (1) The level of density is too high to be explained by water—the temperature factor was refined to unreasonably low values ($<2 \text{ \AA}^2$) when water molecules were assigned but was appropriate when Ca^{2+} were assigned ($\sim 40 \text{ \AA}^2$, similar to those of coordinating atoms). (2) At least six coordinating oxygen atoms are located at a distance of 2.2–2.6 $\text{\AA}$ from the centre of each site; this is unusual for a water molecule. (3) The valence values calculated from the geometry of coordinating atoms are 1.95 (I) and 2.15 (II), which qualify as high-affinity Ca^{2+} -binding sites¹⁸. (4) Omit difference maps showed the presence of a much smaller peak 2.4 $\text{\AA}$ away from the main one (Fig. 4, I). This peak had a density appropriate to a water molecule, as judged by the refined temperature factor.

These two sites are at similar heights with respect to the

membrane (Fig. 1) and are 5.7 $\text{\AA}$ apart (Fig. 4). Following previous proposals, they are termed as sites I and II³. Site I is located in the space between the M5 and M6 helices with contribution from M8 at a rather distal position. The side-chain oxygen atoms of Asn 768, Glu 771 (M5), Thr 799, Asp 800 (M6) and Glu 908 (M8) contribute to this site, in agreement with mutational studies⁶. Disruption of helix conformation around Asp 800 and Gly 801 allows both Thr 799 and Asp 800 on M6 to contribute. All of the side-chain oxygen atoms, except for that of Glu 771, are arranged on roughly the same plane; that of Glu 771 coordinates from underneath (that is, the luminal side of) the plane (Fig. 4b). In contrast, site II is formed almost 'on' helix M4 (Fig. 3a) by the main-chain carbonyl oxygen atoms of Val 304, Ala 305 and Ile 307 (M4) and the side-chain oxygen atoms of Asn 796, Asp 800 (M6) and Glu 309 (M4). Obviously, this kind of coordination geometry is only possible because helix M4 is unwound between Ile 307 and Gly 310. The PEG motif in this unwound region has been recognized as a key motif²; the glutamate residue is replaced by cysteine or histidine in ATPases that transport heavy metals².

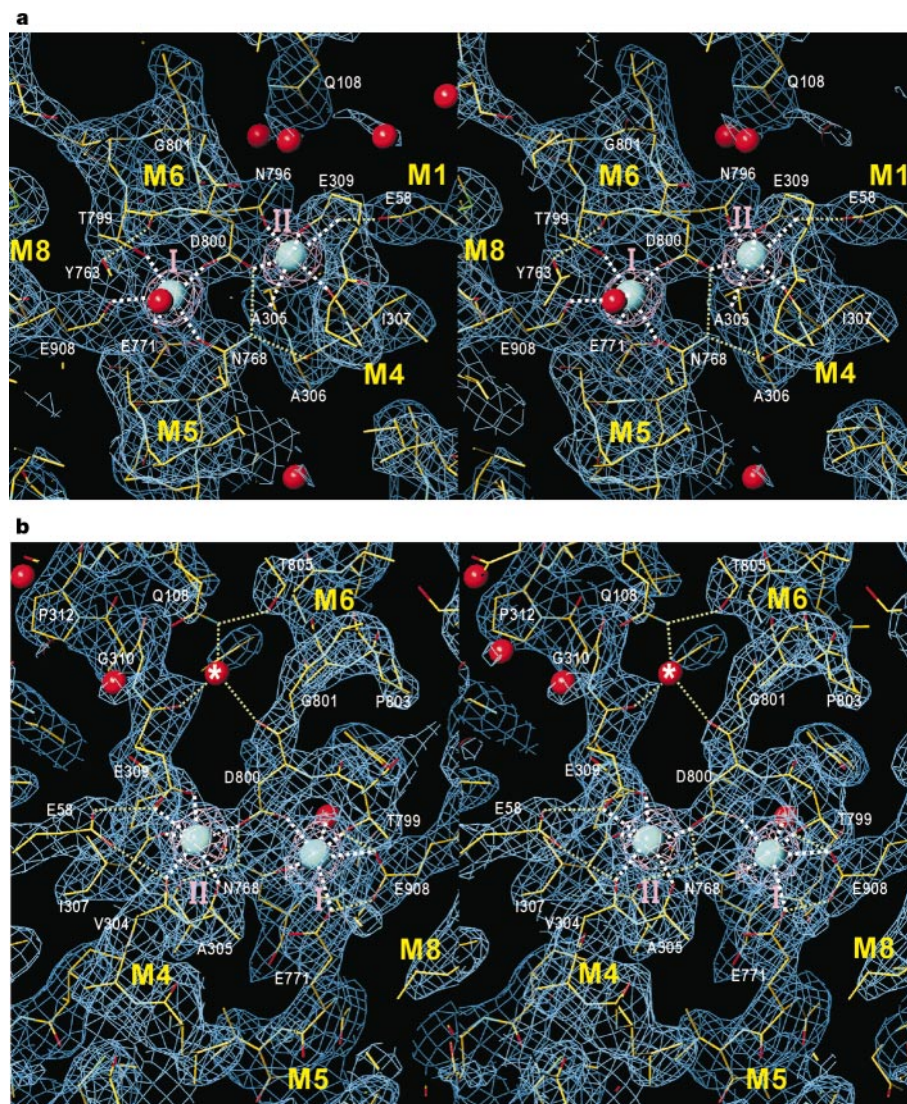


Figure 4 Details of the transmembrane Ca^{2+} -binding sites. The refined model is superimposed with a $2|F_0| - |F_c|$ composite-omit map prepared with CNS (blue meshes, contoured at 1.5σ). The meshes in pink show omit $|F_c| - |F_c|$ map for Ca^{2+} and a bound water (cut-off at 3σ). Blue spheres represent Ca^{2+} , and red spheres water molecules. Viewed roughly normal to the membrane from the cytoplasmic side (**a**) and parallel to the

membrane (**b**) in stereo. The coordinations of oxygen atoms to Ca^{2+} are indicated by white dotted lines, and possible hydrogen bonds stabilizing the coordination geometry by green dashed lines. The viewing direction in **b** is also specified in Fig. 3a. Water molecule nearest to the Ca^{2+} -binding site is marked (asterisk). Note the unwinding of helices M4 and M6 and rows of exposed carbonyl oxygen atoms along these helices in **b**.

These two sites are stabilized by hydrogen-bond networks between the coordinating residues (for example, between Ala 306 and Asn 768) and between residues on other helices (for example, between Val 304 and Glu 309 on M4 and Glu 58 on M1, see Fig. 4). These hydrogen-bond networks must be important for the cooperative binding of two Ca^{2+} ions (ref. 19). As Asp 800 is coordinated to both Ca^{2+} atoms, the correct positioning of this residue must be very important. Furthermore, this residue is located in the unwound region between two helices (that is, the luminal and cytoplasmic halves of M6) and the carbonyl oxygen atom is highly exposed (Fig. 4b), implying that the position of this key residue is under the control of other residues. This unwound conformation agrees well with an NMR study of isolated M6 polypeptide in detergent micelle²⁰.

An important question is: how do Ca^{2+} ions reach the transmembrane binding sites? Owing to the lack of large vestibules similar to those found in ion channels²¹, our structure does not provide a definite answer. One candidate for Ca^{2+} entry is the area surrounded by M2, M4 and M6 (Fig. 3c). It is a cavity with a rather wide opening and is clearly water accessible (seven water molecules are seen in Fig. 3c). In the upper part of this cavity are located Gln 108 and Asn 111, a critical residue for ATPase activity²². Particularly interesting are the rows of exposed oxygen atoms formed by the unwound part of M4 (Pro 312 to Glu 309) and of M6 (Gly 801 and Asp 800). These rows of main-chain carbonyl oxygen atoms point towards the cytoplasm (Fig. 4b), providing a hydrophilic pathway leading to the Ca^{2+} -binding sites. The rows constrict near the Ca^{2+} -binding sites, trapping a water molecule (Figs 3c and 4b, asterisk); here, the oxygen atoms are arranged so that nearly ideal hydrogen-bonding geometry is provided for this water molecule. Such a geometry must be required for displacing water molecules²¹ that are bound to Ca^{2+} much more tightly than to monovalent cations. The outlet of Ca^{2+} is likely to be located in the area surrounded by M3–M5 (Fig. 3a, circle); here too, a ring of oxygen atoms with bound water molecules is formed.

Thus, the unwinding of transmembrane helices appears to be required for two purposes: first, to realize efficient coordination geometry; and second, to provide rows of oxygen atoms for guiding Ca^{2+} to the binding site and removing water molecules at the same time.

Function and structure of three cytoplasmic domains

As described above, the cytoplasmic headpiece consists of three well separated domains. The central part is occupied by domain P, which contains the residue of phosphorylation (Asp 351). This domain is composed of two parts widely separated in the amino-acid sequence. The N-terminal part (roughly between Asn 330 and Asn 359) is connected to M4 and the C-terminal part (roughly Lys 605 to Asp 737) to M5. These two parts are assembled into a seven-stranded parallel β -sheet with eight short associated helices, together forming a typical Rossman fold (Fig. 5a). The residue of phosphorylation, Asp 351, is situated in the C-terminal end of the central β -strand (Fig. 5a), as is usually the case with nucleotide-binding proteins with a Rossman fold. Around this, the residues critical for ATP hydrolysis are clustered (Fig. 5a; see refs 2 and 4 for reviews) and form a highly negatively charged surface that is accessible to solvent (Fig. 6b).

In our structure, the folding pattern and the locations of the critical amino acids are essentially identical to those of the core domain of L-2-haloacid dehalogenase²³ (HAD; Fig. 5b), as predicted by sequence homology⁸. Using this homology, we can propose roles to some extent for the critical amino acids.

Domain N is the largest of the three cytoplasmic domains and has $M_r \approx 27\text{K}$. This domain is formed by the residues inserted (roughly Gln 360–Arg 604) between the two regions constituting domain P and corresponds to the sub-domain in HAD. It comprises a seven-stranded antiparallel β -sheet with two helix bundles sandwiching it (Fig. 2; see also Fig. 7b for a top view). A function of this domain became clear when we directly visualized bound nucleotide. To do this, we soaked the crystals in 0.5 mM TNP-AMP solution and identified the nucleotide's position by difference Fourier methods (Table 1). In the difference Fourier map (Fig. 6a), TNP-AMP is located unambiguously near Phe 487, which has been identified as a critical residue for ATP binding²⁴. Well characterized residues such as Lys 515 and Lys 492 are located nearby (Fig. 6a). Lys 515, thought for a long time to be at the 'mouth' of the ATP-binding pocket²⁵, is actually located in the depth of the binding pocket (Fig. 6b). The close proximity of Lys 515 to the adenosine moiety ($\sim 5\text{\AA}$, Fig. 6a) explains the inhibition of high-affinity ATP binding by chemical modification of this residue with fluorescein isothiocyanate²⁵. Lys 492, labelled with 8-azido-TNP-AMP/ATP²⁶

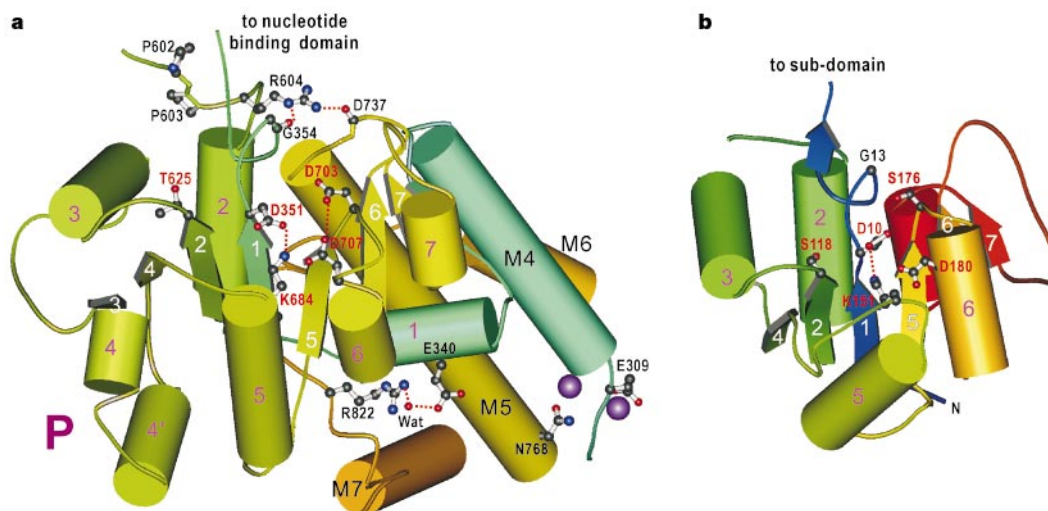


Figure 5 Architecture of the phosphorylation domain (P in Fig. 2) of Ca^{2+} -ATPase (**a**) and the catalytic core domain of L-2-haloacid dehalogenase (HAD²³, **b**). Critical amino-acid residues found in HAD are all preserved and indicated in **a**. D351 is the phosphorylation site in Ca^{2+} -ATPase (**a**), corresponding to D10 in HAD (**b**). Three residues (Thr 625 to Asp 627) forming the loop connecting β -strand 2 and α -helix 3 are all critical for ATPase

activity⁴⁸. Lys 684 is labelled by ATP pyridoxal²⁷ and Asp 703 and Asp 707 are labelled by CIRATP with Na^+K^+ -ATPase⁴⁹. Colour coding is as in Fig. 2. Numbering of the secondary structures in **b** is made consistent with **a** and different from the original²³. Red dashed lines show possible hydrogen bonds. Figure prepared with MOLSCRIPT⁴⁷.

and other ATP-derivatives²⁷ is located in a short loop connecting the β -strands (Lys 492 is marked in Fig. 2; also see Fig. 7b for a top view). This binding pocket is positively charged, in marked contrast to the region around the phosphorylation site (Fig. 6b).

Domain A is the smallest of the three domains and has $M_r \approx 16$ K. It comprises roughly the 110 residues between M2 and M3 (previously called the transduction domain⁵ or β -strand domain), which form a distorted jelly roll structure, and the 40 residues of the N-terminal part of the enzyme, which form two short helices. This domain is connected to the transmembrane region by long loops and is nearly isolated. The susceptibility of the loops to proteinase K attack is affected by Ca^{2+} (ref. 28). The well conserved TGES motif (starting at Thr 181) is located at an edge of this domain (E183 is shown in Fig. 2). Arg 198, trypsin digestion at which is enhanced by Ca^{2+} and blocked by phosphorylation²⁹, is located at the outermost loop (T2 in Fig. 2). All these biochemical data suggest that this domain moves substantially during active transport.

Expected structure in the absence of Ca^{2+}

As mentioned above, the cytoplasmic headpiece has a compact appearance in the maps derived from tubular crystals^{12,13}. The most likely reason for this difference from the present structure is the crystallization conditions: the tubular crystals are formed in the absence of Ca^{2+} and in the presence of decavanadate, whereas our three-dimensional crystals are formed in the presence of 10 mM Ca^{2+} . To examine whether the difference could be explained by domain motions, we fitted the atomic model to a low-resolution map (8 Å) from tubular crystals¹³. This was done first by cutting out the cytoplasmic domains (A, N and P) and then by searching for the best matching positions individually (Fig. 7). The ridges of all surface helices match the density map (blue net) very well. Furthermore, a high-density peak (small sphere in Fig. 7), probably a vanadate oligomer, is found in close proximity to Asp 351. Closer examination shows that a few hydrogen bonds can be readily made at the domain interface.

The fitting of these three cytoplasmic domains leaves one spherical lobe of high density (Fig. 7a, large sphere) unidentified. Its density and size indicate that it is probably decavanadate ($\text{V}_{10}\text{O}_{28}^{6-}$). This lobe is located at the side of the highly positively

charged groove formed by domains N and P (Fig. 6b). The groove is narrower than in the present structure (by $\sim 20^\circ$) and decavanadate fits in the lobe (Fig. 7a), like a ball in jaws preventing the jaws from closing. Arg 489, Lys 492 (in the red loop in domain N in Fig. 7) and Arg 678 (see Fig. 6b) are expected to participate in the binding of this large anion. This location of decavanadate will explain its inhibitory effect on ATPase activity.

To realize this arrangement, all the cytoplasmic domains have to move with respect to the M4 and M5 helices, which do not fit the density map in Fig. 7 and need to be inclined by $\sim 20^\circ$. Domain A undergoes the largest movement of the three (nearly a 90° rotation). The TGES loop starting at Thr 181 (Fig. 7, red loop in domain A) meets the TGD motif starting at Thr 625 (Fig. 7, red loop in domain P; see also Fig. 5a). The critical importance of these two motifs has been well documented². The counterpart of Thr 625 in HAD is Ser 118 (Fig. 5b), which is thought to be the binding residue of the substrate carboxyl group³⁰. The gathering of these loops in the E2 or E2P state has been demonstrated with $\text{Na}^+\text{K}^+\text{-ATPase}$ ³¹.

The large movement of domain A requires changes in the orientations of the M1–M3 helices, consistent with the changes in sensitivity to proteases as mentioned previously²⁸. Domain N has changed its orientation by $\sim 20^\circ$ with respect to domain P, resulting in a similar orientation—when projected along the b axis—to that observed with three-dimensional microcrystals by electron microscopy¹⁴. Despite the poor agreement of helices M4 and M5, the loop connecting helices M6 and M7 (L67) fits well with the density map (Fig. 7a), indicating that it follows domain P. But its position with respect to the membrane helices is different, being much higher above the membrane surface than in the presence of Ca^{2+} . This feature suggests that L67 is important for concerted domain motions.

Role of the loop connecting helices M6 and M7

This loop (L67) runs along the bottom of domain P and in a position that can mediate interactions of domain P and the transmembrane domain (Fig. 8). The P2 helix will link the movements of M5 and the DPPR region, and the P1 helix will couple those of M4 and the central β -strand on which Asp 351 is located (also see Fig. 5a). Thus, this loop interacts directly or indirectly with most of the key parts of the enzyme. If domain P moves or the upper

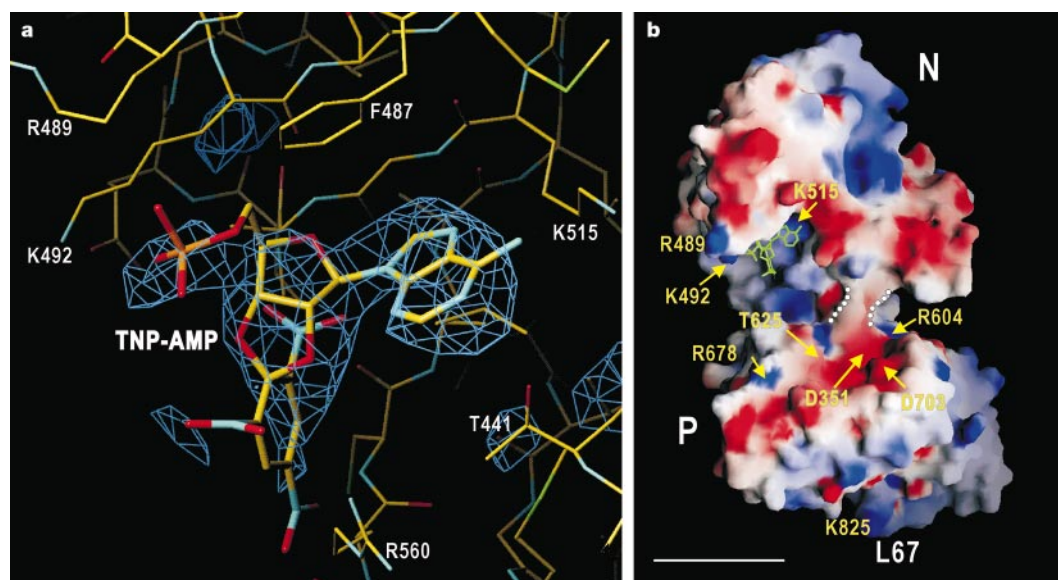


Figure 6 Binding of TNP-AMP to Ca^{2+} -ATPase. Difference Fourier map showing the electron density of bound TNP-AMP (**a**) and a surface potential map (red, negatively charged; blue, positively charged) of domains N and P prepared with Grasp⁵⁰ (**b**). The final model of the protein part (refined without TNP-AMP) and a rough model of TNP-AMP are superimposed (**a**). The map was calculated at 4.0 Å resolution and contoured at 2.5σ (**a**).

In **b**, a stick model of TNP-AMP is placed in the position defined in **a**. White circles identify the bridging region between domains N and P. Note that the distances between the TNP-AMP and the phosphorylation site (D351), and between K492 and R604, which are crosslinked by glutaraldehyde, are more than 25 Å. The direction of view in **b** is indicated in Fig. 2. Scale bar, 20 Å.

part of helix M5 or M4 moves, as suggested by the comparison with the map from tubular crystals, the movement will be transmitted to M3 and M6. Because L67 is directly connected to M6, changes induced in M6 by the binding of Ca^{2+} will be transmitted to domain P through L67 and the P1 or P2 helix.

The comparison with the map from tubular crystals also suggests a conformation change in L67. In this regard, the results of scanning cysteine mutagenesis are interesting. The pair of residues that formed disulphide bridges in the absence of Ca^{2+} were compatible with our atomic model for the luminal half but not for the cytoplasmic half of helix M6³². These results indicate that the cytoplasmic half of M6 may undergo rotation or winding/unwinding during Ca^{2+} transport in concert with the movement of L67. In fact, the geometry of this region is far from an ideal helix and is not recognized as such, depending on the classification program.

Thus, it is likely that this loop is important in coordinating different parts of the enzyme. In this way it may affect Ca^{2+} binding and ATP hydrolysis³³. Asp 813 and Asp 818 were postulated to be the initial Ca^{2+} -binding sites³³, but in this configuration they are separated by ~ 10 Å (Fig. 2) and are unable to form a Ca^{2+} -binding site.

Large domain motion during ATP-hydrolysis

The phosphorylation site (Asp 351) is more than 25 Å away from the bound nucleotide (Fig. 6b). This means that domain closure must occur during ATP hydrolysis. Also, the two residues crosslinked by glutaraldehyde (Lys 492 and Arg 678; ref. 34) are separated by more than 25 Å (Fig. 6b). ATP pyridoxal labels both Lys 492 and Lys 684 in the absence of Ca^{2+} but only Lys 684 (Fig. 5a) in the presence of Ca^{2+} (ref. 27). Fluorescence from TNP-AMP increases during the catalytic cycle and on phosphorylation by P_i (ref. 35). Phosphorescence spectroscopy (of erythrosin isothiocyanate attached to Lys 464 in domain N) showed large domain motions induced by the binding of Ca^{2+} (ref. 36). All these data indicate that domain N is mobile in the presence of Ca^{2+} but fixed in its absence, and that domain N comes close to domain P in the phosphorylated states by thermal fluctuation and/or by conformation changes induced by

ATP-binding. Stabilization of the 'occluded' state by chromium-ATP⁴ is likely to be a result of crosslinking of domains N and P by chromium-ATP.

Large domain motions require the presence of a hinge region. Domain N can be isolated by several proteases³⁷, suggesting that the region around Ser 350 and Gly 609 is indeed flexible. In fact, the connection between domains N and P appears to be extremely tenuous (Figs 2 and 6b), with a well conserved Gly 354 and the invariant DPPR motif² (Fig. 5a). Furthermore, the hydrogen-bond network involving Gly 354, Arg 604 and Asp 737 to Asn 739 seems to link the movement of this hinge region to that of M5, thereby effectively transmitting the phosphorylation signal to the Ca^{2+} -binding sites.

How does Ca^{2+} binding to the high-affinity sites induce large domain motions? It is tempting to postulate that domain A works as an actuator or an anchor for the movement of domain N. As already described, the susceptibility to proteases of the loops connecting domain A to the transmembrane helices changes depending on Ca^{2+} (refs 28, 29), consistent with the idea that Ca^{2+} binding releases domain A. This movement must be made by helices M1–M3, which have only a few hydrogen bonds with other parts of the enzyme. For example, M3 is hydrogen-bonded to L67 near the cytoplasmic end (Fig. 8) and only with M4 at either end. Exactly the same applies to M2. M1 has hydrogen bonds with M4 only, and the residues involved are Glu 309 (Fig. 4) and Lys 297, which are both critical residues⁶. Therefore, Ca^{2+} binding probably stabilizes these hydrogen bonds while destabilizing others, and moves M1–M3 to release domain A. In this regard, it is particularly interesting that thapsigargin, a very potent inhibitor³⁸, is supposed to bind to the region where M3, M4 and L67 gather (around Phe 256; ref. 17; Fig. 8). This must be a very strategic point for fixing them in one configuration. As already noted, L67 may work as a linker between different domains to elicit large-scale domain motions for transporting ions against a concentration gradient.

Conclusion

We have described the atomic model of Ca^{2+} -ATPase with bound

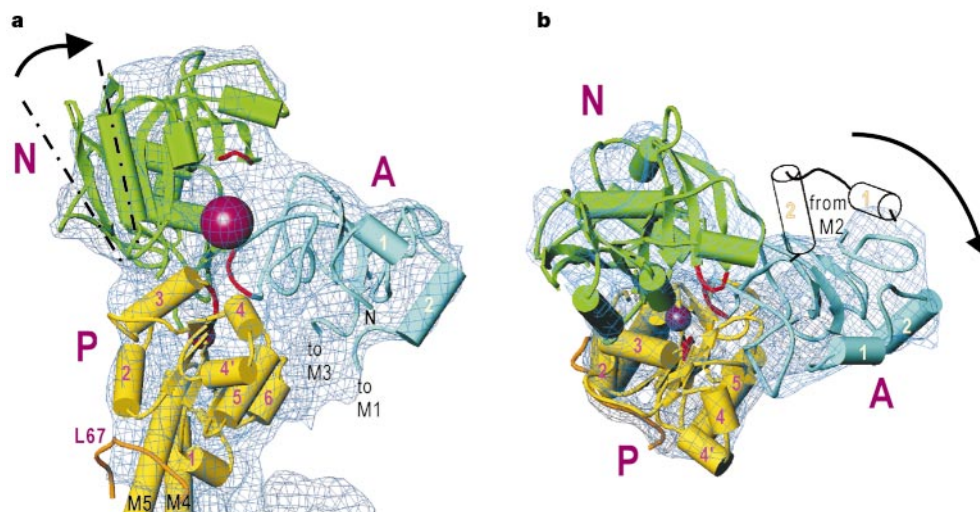


Figure 7 Fitting the atomic model obtained for the Ca^{2+} -bound state (arrows and cylinders) to an 8 Å resolution map (blue net) obtained from tubular crystals formed in the absence of Ca^{2+} and presence of decavanadate and dansyl-thapsigargin¹³. Fitting was done by rigid body motions of three cytoplasmic domains (A, N and P). Large sphere represents a decavanadate ion ($\text{V}_{10}\text{O}_{28}^{6-}$). Small sphere represents the high-density peak around Asp 351; in **a**, it is located between helices P3 and P4'. Red loops correspond to the TGES motif (starting at Thr 181) in domain A (cyan), the RDRK motif (starting at Arg 489) in domain N (green) and the TGD motif (starting at Thr 625) in domain P (yellow). Overall orientation of the molecule in **a** is the same as in Fig. 2a. In **b**, it is viewed from the

cytoplasmic side normal to the membrane. Two dashed-dotted lines in **a** show the orientation of the longest helix in domain N to approximate the movement of domain N (the direction indicated by the arrow). Open cylinders are drawn in **b** for the same purpose (showing the positions of A1 and A2 helices when whole cytoplasmic domains moved as a rigid body to the best fitting position for domain P). The orientation of domain P with respect to M4 and M5 is different from that in 10 mM Ca^{2+} . Hence the fitting of M4 and M5 is poor but can be much improved by inclining them by $\sim 20^\circ$. The residues included for fitting are 1–42 and 125–233 (A), 359–602 (N), and 330–357 and 605–740 (P).

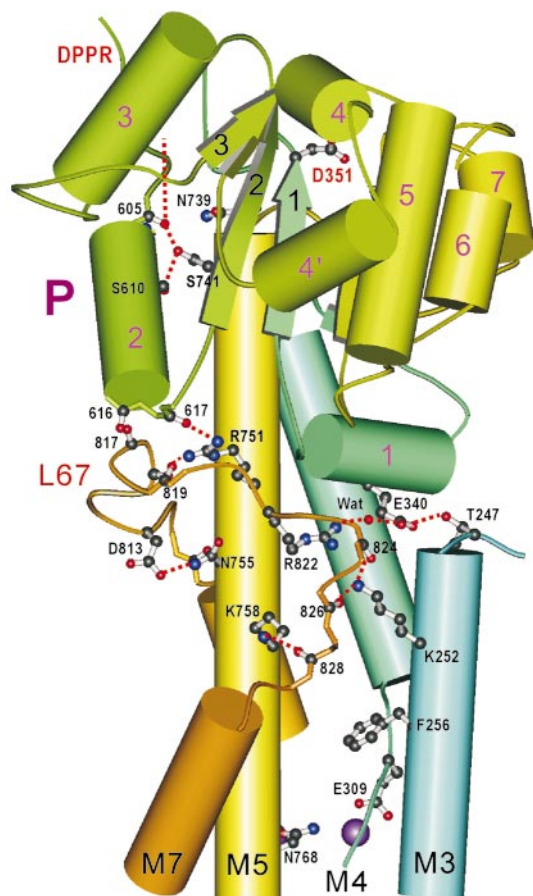


Figure 8 Coupling of the phosphorylation domain and the transmembrane domain by the loop (L67) connecting helices M6 and M7. L67 runs along the bottom of domain P and seems to be important in transmitting signal between the phosphorylation site and the Ca^{2+} -binding sites, which are separated by ~ 50 Å. As depicted, there are extensive hydrogen bonds (blue dotted lines) between this loop and various helices, as well as van der Waals contacts (not shown). Colour coding is as in Fig. 2. F256 is involved in thapsigargin binding¹⁷.

Ca^{2+} on the basis of X-ray diffraction data at 2.6 Å resolution. Our model explains many more experimental results than we can describe here. We have shown that the cytoplasmic region consists of three well separated domains and have proposed new names for them: domain A may work as an actuator or an anchor for domain N, which binds nucleotides; and domain P contains Asp 351, the residue of phosphorylation. The finding that the binding site for the adenosine moiety and the phosphorylation site are located on different domains will settle much of the controversy on the ATP-binding site(s)⁴. We also have described how two Ca^{2+} ions are bound to their transmembrane binding sites and have proposed a possible pathway for Ca^{2+} . We have fitted the atomic model to a low-resolution map of the enzyme without Ca^{2+} but with decavanadate bound, and shown approximate configurations of different domains in this state. This comparison has indicated a mechanism for how the cytoplasmic domains communicate with the helices that form the Ca^{2+} -binding site, which is more than 50 Å away (the distance between the ATP-binding pocket and the bound Ca^{2+} ions is ~ 80 Å). □

Methods

Ca^{2+} -ATPase was prepared from rabbit hind leg muscle and purified using affinity column chromatography¹¹ after solubilization using octaethyleneglycol mono-n-dodecylether (C_{12}E_8). Crystals were grown by dialysing the mixture of purified protein and phospholipid (phosphatidylcholine) against crystallization buffer³⁹ containing 0.8 M sodium butyrate, 2.75 M glycerol, 10 mM CaCl_2 , 3 mM MgCl_2 , 2.5 mM NaN_3 , 0.2 mM dithiothreitol and 20 mM MES, pH 6.1. After one month of incubation, the crystals were grown

to typically $100 \times 500 \times <20$ μm. The crystals belonged to space group C2 with unit cell parameters of $a = 166.0$ Å, $b = 64.4$ Å, $c = 147.1$ Å, $\beta = 98.0^\circ$ (at 100 K), and contained one protein molecule in the asymmetric unit. The thickness of the crystals used for diffraction experiments was occasionally measured by freeze-substitution and sectioning of the crystals. The crystals were picked up using nylon loops and flash-frozen in cold nitrogen gas from a cryostat (Rigaku). Heavy-atom derivatives were prepared by soaking for 1 h. Phase information of the (h k 0) plane derived from electron micrographs of the thinner crystals was very useful in efficient screening of derivatives with a laboratory X-ray source. Final data were all collected at SPring-8 (beam lines BL41XU (ref. 40) and BL44B2 (ref. 41)) using either Rigaku R-Axis IV or Mar CCD 165 detectors.

Data taken with R-Axis IV were processed with Denzo and Scalepack⁴² and those with the CCD detectors were processed with MOSFLM⁴³ and Scala⁴⁴. Merging of different data sets for native data was done using Scalepack. Subsequent data processing was carried out using CCP4 program suites⁴⁴. Experimental phases were first calculated by Mlphare⁴⁴ to 2.8 Å resolution derived from five platinum and three lanthanide derivatives (Table 1). The mean figure of merit for acentric reflections was 0.558 (20.0–2.8 Å) and 0.294 in the last resolution shell (3.17–2.80 Å). The map refined by density modification with Solomon⁴⁴ (Fig. 3a, b) was used for the initial modelling with Turbo-Prodo (BioGraphics). For structure refinement CNS⁴⁵ was used. The refined temperature factors were 60.0 Å² for protein atoms, 48.7 Å² for bound water molecules and 41.3 Å² for Ca^{2+} . The assignment of secondary structures was done using the program DSSP⁴⁶. Fitting of the atomic model to the low resolution map by electron microscopy was first done manually using Turbo-Prodo and then refined by calculating the correlation function with magnification correction of the electron microscopy map.

Received 9 March; accepted 4 May 2000.

1. Ebashi, S. & Lipman, F. Adenosine triphosphate-linked concentration of calcium ions in a particulate fraction of rabbit muscle. *J. Cell. Biol.* **14**, 389–400 (1962).
2. Møller, J. V., Juul, B. & le Maire, M. Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochim. Biophys. Acta* **1286**, 1–51 (1996).
3. MacLennan, D. H., Rice, W. J. & Green, N. M. The mechanism of Ca^{2+} transport by sarco(endo)-plasmic reticulum Ca^{2+} -ATPases. *J. Biol. Chem.* **272**, 28815–28818 (1997).
4. McIntosh, D. B. The ATP binding sites of P-type ion transport ATPases. *Adv. Mol. Cell. Biol.* **23A**, 33–99 (1998).
5. MacLennan, D. H., Brandl, C. J., Korczak, B. & Green, N. M. Amino-acid sequence of a Ca^{2+} + Mg^{2+} -dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence. *Nature* **316**, 696–700 (1985).
6. Clarke, D. M., Loo, T. W., Inesi, G. & MacLennan, D. H. Location of high affinity Ca^{2+} -binding sites within the predicted transmembrane domain of the sarcoplasmic reticulum Ca^{2+} -ATPase. *Nature* **339**, 476–478 (1989).
7. Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**, 430–434 (1990).
8. Aravind, L., Galperin, M. Y. & Koonin, E. V. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.* **23**, 127–129 (1998).
9. Dux, L. & Martonosi, A. Two-dimensional arrays of proteins in sarcoplasmic reticulum and purified Ca^{2+} -ATPase vesicles treated with vanadate. *J. Biol. Chem.* **258**, 2599–2603 (1983).
10. Dux, L., Pikula, S., Mullner, N. & Martonosi, A. Crystallization of Ca^{2+} -ATPase in detergent-solubilized sarcoplasmic reticulum. *J. Biol. Chem.* **262**, 6439–6442 (1987).
11. Stokes, D. L. & Green, N. M. Three-dimensional crystals of CaATPase from sarcoplasmic reticulum. Symmetry and molecular packing. *Biophys. J.* **57**, 1–14 (1990).
12. Toyoshima, C., Sasabe, H. & Stokes, D. L. Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane. *Nature* **362**, 467–471 (1993).
13. Zhang, P., Toyoshima, C., Yonekura, K., Green, N. M. & Stokes, D. L. Structure of the calcium pump from sarcoplasmic reticulum at 8-Å resolution. *Nature* **392**, 835–839 (1998).
14. Ogawa, H., Stokes, D. L., Sasabe, H. & Toyoshima, C. Structure of the Ca^{2+} pump of sarcoplasmic reticulum: a view along the lipid bilayer at 9-Å resolution. *Biophys. J.* **75**, 41–52 (1998).
15. Brandl, C. J., deLeon, S., Martin, D. R. & MacLennan, D. H. Adult forms of the Ca^{2+} -ATPase of sarcoplasmic reticulum. Expression in developing skeletal muscle. *J. Biol. Chem.* **262**, 3768–3774 (1987).
16. Toyofuku, T., Kurzydowski, K., Tada, M. & MacLennan, D. H. Amino acids Lys-Asp-Asp-Lys-Pro-Val⁴⁰² in the Ca^{2+} -ATPase of cardiac sarcoplasmic reticulum are critical for functional association with phospholamban. *J. Biol. Chem.* **269**, 22929–22932 (1994).
17. Yu, M. *et al.* Specific substitutions at amino acid 256 of the sarcoplasmic/endoplasmic reticulum Ca^{2+} transport ATPase mediate resistance to thapsigargin in thapsigargin-resistant hamster cells. *J. Biol. Chem.* **273**, 3542–3546 (1998).
18. Nayal, M. & Di Cera, E. Predicting Ca^{2+} -binding sites in proteins. *Proc. Natl Acad. Sci. USA* **91**, 817–821 (1994).
19. Inesi, G., Kurzmack, M., Coan, C. & Lewis, D. E. Cooperative calcium binding and ATPase activation in sarcoplasmic reticulum vesicles. *J. Biol. Chem.* **255**, 3025–3031 (1980).
20. Soulié, S. *et al.* NMR conformational study of the sixth transmembrane segment of sarcoplasmic reticulum Ca^{2+} -ATPase. *Biochemistry* **38**, 5813–5821 (1999).
21. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^{+} conduction and selectivity. *Science* **280**, 69–77 (1998).
22. Clarke, D. M. *et al.* Functional consequences of glutamate, aspartate, glutamine, and asparagine mutations in the stalk sector of the Ca^{2+} -ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* **264**, 11246–11251 (1989).
23. Hisano, T. *et al.* Crystal structure of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL. An α/β hydrolase structure that is different from the α/β hydrolase fold. *J. Biol. Chem.* **271**, 20322–20330 (1996).
24. McIntosh, D. B., Woolley, D. G., Vilsen, B. & Andersen, J. P. Mutagenesis of segment 487 Phe-Ser-Arg-Asp-Arg-Lys⁴⁸⁷ of sarcoplasmic reticulum Ca^{2+} -ATPase produces pumps defective in ATP binding. *J. Biol. Chem.* **271**, 25778–25789 (1996).
25. Pick, U. Interaction of fluorescein isothiocyanate with nucleotide-binding sites of the Ca-ATPase from sarcoplasmic reticulum. *Eur. J. Biochem.* **121**, 187–195 (1981).

26. McIntosh, D. B., Woolley, D. G. & Berman, M. C. 2',3'-O-(2,4,6-trinitrophenyl)-8-azido-AMP and-ATP photolabel Lys-492 at the active site of sarcoplasmic reticulum Ca^{2+} -ATPase. *J. Biol. Chem.* **267**, 5301–5309 (1992).
27. Yamamoto, H., Imamura, Y., Tagaya, M., Fukui, T. & Kawakita, M. Ca^{2+} -dependent conformational change of the ATP-binding site of Ca^{2+} -transporting ATPase of sarcoplasmic reticulum as revealed by an alteration of the target-site specificity of adenosine triphosphopyridoxal. *J. Biochem. (Tokyo)* **106**, 1121–1125 (1989).
28. Juul, B. *et al.* Do transmembrane segments in proteolyzed sarcoplasmic reticulum Ca^{2+} -ATPase retain their functional Ca^{2+} binding properties after removal of cytoplasmic fragments by proteinase K? *J. Biol. Chem.* **270**, 20123–20134 (1995).
29. Andersen, J. P., Vilsen, B., Collins, J. H. & Jørgensen, P. L. Localization of E1–E2 conformational transitions of sarcoplasmic reticulum Ca-ATPase by tryptic cleavage and hydrophobic labeling. *J. Membr. Biol.* **93**, 85–92 (1986).
30. Li, Y. F. *et al.* Crystal structures of reaction intermediates of L-2-haloacido dehalogenase and implications for the reaction mechanism. *J. Biol. Chem.* **273**, 15035–15044 (1998).
31. Goldshleger, R. & Karlsh, S. J. D. Fe-catalyzed cleavage of the alpha subunit of Na/K-ATPase: evidence for conformation-sensitive interactions between cytoplasmic domains. *Proc. Natl Acad. Sci. USA* **94**, 9596–9601 (1997).
32. Rice, W. J., Green, N. M. & MacLennan, D. H. Site-directed disulfide mapping of helices M4 and M6 in the Ca^{2+} binding domain of SERCA1a, the Ca^{2+} ATPase of fast twitch skeletal muscle sarcoplasmic reticulum. *J. Biol. Chem.* **272**, 31412–31419 (1997).
33. Falson, P. *et al.* The cytoplasmic loop between putative transmembrane segments 6 and 7 in sarcoplasmic reticulum Ca^{2+} -ATPase binds Ca^{2+} and is functionally important. *J. Biol. Chem.* **272**, 17258–17262 (1997).
34. McIntosh, D. B. Glutaraldehyde cross-links Lys-492 and Arg-678 at the active site of sarcoplasmic reticulum Ca^{2+} -ATPase. *J. Biol. Chem.* **267**, 22328–22335 (1992).
35. Nakamoto, R. K. & Inesi, G. Studies of the interactions of 2',3'-O-(2,4,6-trinitrocyclohexyldienylidene)adenosine nucleotides with the sarcoplasmic reticulum (Ca^{2+} + Mg^{2+})-ATPase active site. *J. Biol. Chem.* **259**, 2961–2970 (1984).
36. Huang, S. & Squier, T. C. Enhanced rotational dynamics of the phosphorylation domain of the Ca-ATPase upon calcium activation. *Biochemistry* **37**, 18064–18073 (1998).
37. Champeil, P. *et al.* Characterization of a protease-resistant domain of the cytosolic portion of sarcoplasmic reticulum Ca^{2+} -ATPase. *J. Biol. Chem.* **273**, 6619–6631 (1998).
38. Sagara, Y. & Inesi, G. Inhibition of the sarcoplasmic reticulum Ca^{2+} transport ATPase by thapsigargin at subnanomolar concentrations. *J. Biol. Chem.* **266**, 13503–13506 (1991).
39. Misra, M., Taylor, D., Oliver, T. & Taylor, K. Effect of organic anions on the crystallization of the Ca^{2+} -ATPase of muscle sarcoplasmic reticulum. *Biochim. Biophys. Acta* **1077**, 107–118 (1991).
40. Kamiya, N. *et al.* Design of the high energy undulator pilot beamline for macromolecular crystallography at the SPring-8. *Rev. Sci. Instrum.* **66**, 1703–1705 (1995).
41. Adachi, S., Oguchi, T. & Ueki, T. Present status of RIKEN beamline II (BL44B2) for structural biology. *SPring-8 Annual Report* **1**, 239–240 (1996).
42. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–325 (1997).
43. Leslie, A. G. W. MOSFLM- recent changes and future developments. *CCP4 Newsletter on Protein Crystallography* **35**, 18–19 (1998).
44. Collaborative Computational project, No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
45. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
46. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
47. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
48. Clarke, D. M., Loo, T. W. & MacLennan, D. H. Functional consequences of alterations to amino acids located in the nucleotide binding domain of the Ca^{2+} -ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* **265**, 22223–22227 (1990).
49. Ovchinnikov, Y. *et al.* Affinity modification of E1-form of Na^+ , K^+ -ATPase revealed Asp-710 in the catalytic site. *FEBS Lett.* **217**, 111–116 (1987).
50. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Acknowledgements

We thank S. Adachi, N. Kamiya (Riken) and M. Kawamoto (JASRI) for their help in data taking at SPring-8, and K. Tani for computations. We acknowledge that very first crystals for X-ray were made by H. Mukai. This work was supported in part by grants-in-aid from the Ministry of Culture, Education, Science and Sports of Japan and from CRESTO (to C.T. and M.N.) and also by Toray Science Foundation (C.T.). This paper is dedicated to S. Ebashi, a founder of Ca^{2+} -ATPase field and the father of the calcium theory.

Correspondence and requests for materials should be addressed to C.T. (e-mail: ct@iam.u-tokyo.ac.jp). The atomic coordinates have been deposited in the PDB (accession code 1EUL).

Newly synthesized lithium in the interstellar medium

D. C. Knauth*, S. R. Federman*, David L. Lambert† & P. Crane‡

* Department of Physics & Astronomy, University of Toledo, Toledo, Ohio 43606, USA

† Department of Astronomy, University of Texas, Austin, Texas 78712, USA

‡ Department of Physics & Astronomy, Dartmouth College, Hanover, New Hampshire 03755, USA, and NASA Headquarters, Washington DC, 20546, USA

Astronomical observations of elemental and isotopic abundances provide the means to determine the source of elements and to reveal their evolutionary pathways since the formation of the Galaxy some 15 billion years ago. The abundance of lithium is particularly interesting because, although some of it is thought to be primordial, most results from spallation reactions (in which Galactic cosmic rays break apart larger nuclei in the interstellar medium). Spallation reactions are crucial for the production of other light elements¹, such as beryllium and boron, so observations of lithium isotopic abundances can be used to test model predictions^{2–5} for light-element synthesis in general. Here we report observations of ⁷Li and ⁶Li abundances in several interstellar clouds lying in the direction of the star α Persei. We find the abundance ratio ⁷Li/⁶Li to be about 2, which is significantly lower than the average Solar System value of 12.3 (refs 6, 7). An abundance ratio of 2 is clear evidence that the observed lithium must have resulted entirely from spallation, confirming a basic tenet of light-element synthesis^{2–5}. The total lithium abundance, however, is not enhanced as expected.

The observations were acquired with the Harlan J. Smith 2.7-m telescope and the '2dcoudé' spectrograph at McDonald Observatory⁸. We observed ζ Per (B1 Ib; $V = 2.85$; $v \sin i = 59 \text{ km s}^{-1}$) in November 1996 and January 1997 and α Per (B1 III; $V = 3.83$; $v \sin i = 85 \text{ km s}^{-1}$) in January 1998. High resolution spectra ($R \approx 170,000$) in two spectral regions were obtained, at 4,044 Å for K I and 6,707 Å for Li I. Figure 1 displays the spectra, after removal of the undulating stellar continuum by using a low-order polynomial.

Generally there are multiple interstellar clouds on any given line of sight whose existence is revealed by absorption at distinct Doppler velocities. The velocity structure has to be well known before an accurate lithium isotope ratio can be deduced from the spectra^{9,10}. Therefore, the observed directions should have relatively simple velocity structure (that is, one or two interstellar clouds with detectable Li I lines). In addition, data on another species, which has

similar properties to lithium and resides in the same portion of the interstellar cloud, should be obtained for use as a template of the velocity structure. Our method differs slightly from the technique of Lemoine *et al.*^{9,10} who used K I $\lambda 7,699$ as their velocity template. This line of K I is too strong to use as an effective velocity template. Instead, we obtained data on K I $\lambda 4,044$ which is of comparable strength to the lithium lines (see Fig. 1) and is therefore a more appropriate template. (The oscillator strength for $\lambda 4,044$ is about a factor of 50 smaller¹¹.) Moreover, high signal-to-noise, high resolution spectra of the Li doublet at 6,707 Å are required, as the lines are weak and the fine structure splitting for ⁷Li I is comparable to the isotope shift of ⁶Li I (~ 0.160 Å), resulting in a blend of the ⁷Li and ⁶Li lines, as shown in Fig. 1.

The observations were fitted by synthetic profiles where profile parameters were adjusted until a minimum χ^2 was obtained; this method was successfully applied in our determination of the interstellar ¹¹B/¹⁰B ratio^{12,13}. The results from applying the parameters obtained independently from the ⁷Li I and K I lines show agreement in column densities at the 1 σ level. The results of profile fitting are displayed in Table 1, and Table 2 contains the inferred isotope ratios.

Our ⁷Li/⁶Li ratio of 10.6 ± 2.9 for the cloud toward ζ Per is just consistent with a previous determination¹⁴ ($5.5^{+1.3}_{-1.1}$), the difference being in the derived $N(^6\text{Li})$, and agrees with the most recent meteoritic value⁷ (11.86–12.44). As shown by inspection of the line profile, the ⁷Li/⁶Li ratios for α Per's clouds are highly non-solar: ⁷Li/⁶Li = 3.6 ± 0.6 and 1.7 ± 0.3 , for the stronger and weaker cloud, respectively, with the latter being essentially the value (≈ 2) predicted for Li production by cosmic ray-induced spallation. The most recent models^{4,5} use up-to-date cross-sections for spallation reactions and cosmic ray data. Reactions of importance for the Li isotopes are protons and helium nuclei on C, N and O nuclei, and there is a contribution from helium–helium fusion. Considering recent results for the cosmic ray flux, composition and propagation, and incorporating the effects of ionization of interstellar gas, the computed ⁷Li/⁶Li ratios are 1.5–1.7 (ref. 4) and 1.35 ± 0.40 (ref. 5), values equal to the lower of the ratios found for α Per. Our conclusion is that the cloud material has been subjected to a severe flux of energetic particles. Such a flux has not affected ζ Per's cloud. The line of sight toward α Per lies closer than that of ζ Per to a site of star formation, IC 348, where massive stars could supply energetic particles via supernovae or stellar winds. Confirmation is provided by a study of the interstellar OH and HD chemistry of these sight lines that showed the cosmic ray flux toward α Per to be approximately an order of magnitude higher^{15,16}.

An earlier study¹⁰ suggested a low ⁷Li/⁶Li ratio for a cloud toward ζ Oph. Using the stronger K I line at 7,699 Å as a velocity template, isotope ratios of $8.8 \pm 0.8 \pm 1.4$ and $1.4^{+1.2}_{-0.5} \pm 0.6$, where the second uncertainty represents systematic effects, were obtained.

Table 1 Results of profile fitting

Parameter	α Per			ζ Per		
	⁷ Li fit	K fit	Average	⁷ Li fit	K fit	Average
b -value (km s ⁻¹)	2.1	2.1	2.1	1.9	1.4	1.7
	0.6	0.8	0.7			
v_{lsr} (km s ⁻¹)	6.5	6.5	6.5	6.5	6.5	6.5
	3.5	3.3	3.4			
$N(^7\text{Li})$ (10 ⁸ cm ⁻²)	26 ± 2	26 ± 2	26 ± 1	37 ± 2	33 ± 2	35 ± 1
	5.6 ± 0.6	5.3 ± 0.8	5.5 ± 0.5			
$N(^6\text{Li})$ (10 ⁸ cm ⁻²)	7.0 ± 1.5	7.4 ± 1.5	7.2 ± 1.1	3.7 ± 1.4	3.0 ± 1.1	3.3 ± 0.9
	3.2 ± 0.6	3.2 ± 0.8	3.2 ± 0.5			
$N(\text{K I})$ (10 ¹¹ cm ⁻²)	9.4 ± 0.3	9.4 ± 0.4	9.4 ± 0.2	7.6 ± 0.4	7.0 ± 0.4	7.3 ± 0.3
	1.8 ± 0.2	1.9 ± 0.2	1.9 ± 0.1			

The synthesis included hyperfine structure for both ⁶Li and ⁷Li (ref. 22). The hyperfine structure splitting of K I $\lambda 4,044$ was found to be negligible (~ 0.002 mÅ) and was not included in the analysis. The weaker member of the K I doublet is at 4,047 Å and thus is not part of the analysis. For oscillator strengths, we adopted tabulated results for each hyperfine transition in Li I (ref. 23) and for the fine structure line of K I (ref. 11). The profiles of the blue ⁷Li fine structure line and K I line were fitted first to obtain the b -value and velocity of the line in the local standard of rest (v_{lsr}), as well as column density (N) for each species. The b -value, which is dominated by turbulent motion, and velocity were then fixed and used in the fit of the Li I profile. The uncertainties in N were determined from the uncertainties associated with the amount of absorption, while for the b -value and v_{lsr} , the results for the two fits provide a measure of their uncertainties. The final results toward each star are weighted averages determined by the separate fits. The results for component 1 toward α Per are listed above those for component 2. The larger b -value for the ⁷Li fit toward ζ Per arises in part from the November data having a slightly larger instrumental width. For the weak lines examined here, the difference in b -value produces a 10% effect in column density and in the inferred isotope ratio.

One concern with the two-component model for ζ Oph (ref. 10) is that the main component is known to be 2 components separated by 1 km s^{-1} (ref. 17). Another concern is that there is no evidence for molecular absorption from the second component toward ζ Oph, unlike the situation in other directions¹⁷. Our measurements, acquired at higher spectral resolution and analysed with a more appropriate template, offer the first (to our knowledge) clear and unambiguous evidence for freshly synthesized products of cosmic ray spallation.

Derivation of the total interstellar Li/H abundance is not trivial because knowledge of the abundance for Li II, the dominant ion, and mechanisms for depletion onto interstellar grains is necessary. An estimate for Li II abundance from data on Li I requires the electron density, which can be approximated by the amount of C II, the most abundant element providing electrons in neutral interstellar gas. The resulting total Li/H abundances appear in Table 2. Since a Li isotope ratio of about 2 toward o Per suggests newly processed Li, the comparable Li/H abundances are unexpected. The two abundances, however, are very similar to recent interstellar Li results, using the analysis described in the legend to Table 2 for consistency: 13.1×10^{-10} toward ζ Per (ref. 14); 11.0×10^{-10} toward ρ Oph (ref. 9); $(20.6\text{--}24.4) \times 10^{-10}$ toward ζ Oph (refs 10, 14). For further comparison, the Solar System value⁶ is 20×10^{-10} . Since a weighted interstellar average of C II was utilized in determining the electron density n_e toward o Per, there is additional uncertainty in the derived lithium abundance. There also remains the poorly known correction for depletion onto grains.

Another observational measure, the K/Li ratio, may shed light on this inconsistency. This ratio is not dependent on electron density or the strength of the radiation field, and the amounts of depletion

onto grains are similar¹⁸. The inferred K/Li ratios displayed in Table 2 for the clouds toward o Per and ζ Per show no clear trend and are similar to the Solar System ratio⁶ of 66 ± 8 . Again, no indication for fresh Li is evident.

Although the measurements are as yet few in number, a reasonable assumption based on isotopic Li measurements for several clouds, local stars¹⁹ and the Solar System is that ${}^7\text{Li}/{}^6\text{Li} \approx 10$ once characterized o Per's clouds. If gas with such an isotopic ratio is exposed to energetic particles, fresh ${}^7\text{Li}$ and ${}^6\text{Li}$ are synthesized but little lithium is destroyed. Dilution of ${}^7\text{Li}$ in this fashion would necessarily not attain the observed low isotopic ratio until the total lithium abundance had been raised about an order of magnitude. Neither the Li abundance nor the K/Li ratio in o Per's clouds are, however, unusual; Li appears to be less abundant. This suggests that the energetic particles operated on gas highly depleted in lithium: ejecta from massive stars—winds and supernovae—may be the playground for the energetic particles.

Models²⁰ of superbubbles, which contain ionized gas blown by supernovae in star-forming regions, show that energetic particles synthesize about 10^{51} Li atoms. At the end of the superbubble's existence, these Li atoms are diluted with swept-up ambient material at which point the Li/H ratio is 100 times less than observed toward o Per. The clouds toward o Per, if spherical, contain 3×10^{48} Li atoms in three solar masses of gas. If the superbubble is itself roughly spherical, the total Li content inferred from the o Per sight line is close to the predicted 10^{51} atoms. The superbubble scenario may account for our result provided that dilution of its products with ambient gas has not proceeded to completion. Sight lines containing partially mixed products from a superbubble like that toward o Per may be rare.

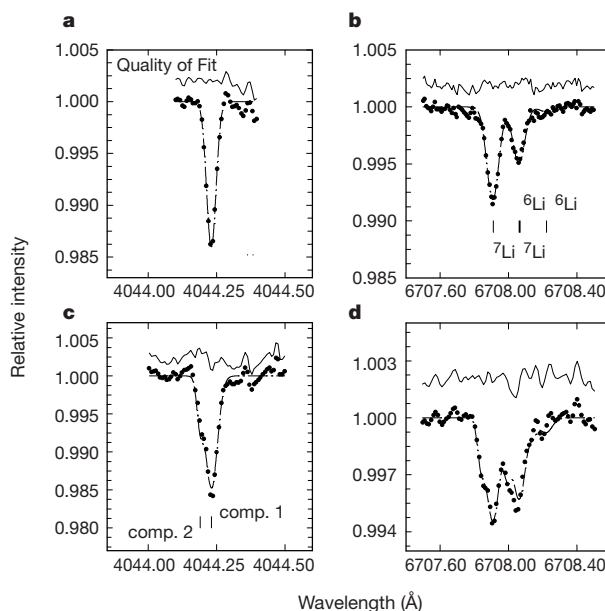


Figure 1 Interstellar spectra of K I and Li I toward ζ and o Per. **a**, K I toward ζ Per; **b**, Li I toward ζ Per; **c**, K I toward o Per; **d**, Li I toward o Per. The spectra were reduced with NOAO SUN/IRAF. After correcting for the bias and removing scattered light, the stellar images were flat-fielded to remove pixel-to-pixel variations in sensitivity. Extracted spectra were produced by summing over pixels perpendicular to the dispersion direction and were placed on a wavelength scale with spectra from a Th–Ar hollow cathode taken throughout the night. Individual spectra were Doppler-corrected and co-added, yielding a signal-to-noise ratio per pixel of approximately 2,500:1 for each stellar spectrum based on the peak-to-peak variation in the continuum. The data points are filled circles. The dot-dashed line is the fit to the line profiles using the K I line for the velocity structure. The solid line gives the difference between data and fit; the offset is 1.002. The vertical lines in the

K I plot for o Per show the positions of the two velocity components; a single velocity component is seen toward ζ Per. The vertical dashes in the Li I plot for ζ Per indicate positions of the ${}^7\text{Li}$ and ${}^6\text{Li}$ fine structure lines. In the absence of ${}^6\text{Li}$, an unsaturated ${}^7\text{Li}$ feature from a single cloud is a doublet with the red component half the strength of the blue component. A ${}^6\text{Li}$ contribution augments the strength of the red component and contributes a weaker redder component. As long as the ${}^7\text{Li}/{}^6\text{Li}$ ratio is not small, the Li I feature is dominated by the 2:1 ratio of the blue to red component of (primarily) ${}^7\text{Li}$. For o Per, however, the red component approaches the strength of the blue component. Since K I $\lambda 4,044$ shows that this enhancement cannot be due to a red-shifted ${}^7\text{Li}$ component, the clouds toward o Per must be rich in ${}^6\text{Li}$.

Table 2 Results for the ${}^7\text{Li}/{}^6\text{Li}$ ratio, the total Li/H abundance and the K/Li abundance ratio

Parameter	α Per	ζ Per
${}^7\text{Li}/{}^6\text{Li}$ ratio	3.6 ± 0.6 1.7 ± 0.3	10.6 ± 2.9
Li/H abundance	$(9.8 \pm 3.5) \times 10^{-10}$	$(12.2 \pm 2.2) \times 10^{-10}$
K/Li abundance ratio	65 ± 4 48 ± 4	42 ± 2

For our estimates of the total Li/H abundance, the electron density (n_e) is obtained from $N(\text{C II})$, the total proton column density [$N(\text{H}) = N(\text{H I}) + 2N(\text{H}_2)$], and the gas density (n). Since no precise $N(\text{C II})$ measurements exist for the direction toward α Per, the weighted mean interstellar ratio²⁴ of $N(\text{C II})/N(\text{H}) = (1.42 \pm 0.13) \times 10^{-4}$ was utilized, yielding $N(\text{C II}) = (2.2 \pm 0.8) \times 10^{17} \text{ cm}^{-2}$. The value of $N(\text{C II})$ is $(1.84 \pm 0.32) \times 10^{17} \text{ cm}^{-2}$ toward ζ Per (ref. 25). The columns $N(\text{H})$ (refs 26, 27) are $(15.2 \pm 5.3) \times 10^{20} \text{ cm}^{-2}$ toward α Per and $(15.8 \pm 4.7) \times 10^{20} \text{ cm}^{-2}$ toward ζ Per. Chemical models for carbon-bearing molecules²⁸ indicate values for n of 800 and 700 cm^{-3} toward α and ζ Per, respectively. The total Li/H abundance is derived from $\{[N({}^7\text{Li}) + N({}^6\text{Li})]/N(\text{H})\} \times [G(\alpha n_e)]$, where a value of 41 was used for $[G(\alpha n_e)]$, the photoionization rate to recombination rate coefficient¹⁵. The ionization corrections for the two clouds toward α Per are assumed to be the same. The K/Li ratios are based on column densities in Table 1 and a value of 9.4 for $[G(\alpha n_e)]$ (ref. 18). For the ${}^7\text{Li}/{}^6\text{Li}$ and K/Li ratios toward α Per, component 1 is the first entry and component 2 is the second.

The presence of deuterium as HD molecules toward α Per (ref. 21) implies that the gas within the superbubble cannot have been totally depleted in Li initially. The D-containing gas may be largely in the cloud for which we find ${}^7\text{Li}/{}^6\text{Li} \approx 4$. Synthesis of D within a superbubble is negligible. The existence of D may warrant consideration of alternative hypotheses, including (1) isotopic fractionation of Li in diffuse clouds and (2) the idea that pristine interstellar gas has ${}^7\text{Li}/{}^6\text{Li} \approx 2$, the value for cosmic ray spallation, and so varying degrees of contamination with ejecta from Li-rich red giants drive the ratio to the higher values seen in ζ Per's clouds and elsewhere. □

Received 18 February; accepted 19 April 2000.

- Reeves, H., Fowler, W. A. & Hoyle, F. Galactic cosmic ray origin of Li, Be and B in stars. *Nature* **226**, 727–729 (1970).
- Meneguzzi, M., Audouze, J. & Reeves, H. The production of the elements Li, Be, B by galactic cosmic rays in space and its relation with stellar observations. *Astron. Astrophys.* **15**, 337–359 (1971).
- Meneguzzi, M. & Reeves, H. Light element production by cosmic rays. *Astron. Astrophys.* **40**, 99–110 (1975).
- Ramaty, R., Kozlovsky, B. & Lingenfelter, R. E. Light isotopes, extinct radioisotopes, and gamma-ray lines from low-energy cosmic-ray interactions. *Astrophys. J.* **456**, 525–540 (1996).
- Lemoine, M., Vangioni-Flam, E. & Cassé, M. Galactic cosmic rays and the evolution of light elements. *Astrophys. J.* **499**, 735–745 (1998).
- Anders, E. & Grevesse, N. Abundances of the elements: meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
- Chaussidon, M. & Robert, F. ${}^7\text{Li}/{}^6\text{Li}$ and ${}^{11}\text{B}/{}^{10}\text{B}$ variations in chondrules from the Semarkona unequilibrated chondrite. *Earth Planet. Sci. Lett.* **164**, 577–589 (1998).
- Tull, R. G., MacQueen, P. J., Sneden, C. & Lambert, D. L. The high resolution cross-dispersed echelle white-pupil spectrometer of the McDonald Observatory 2.7 m telescope. *Publ. Astron. Soc. Pacif.* **107**, 251–264 (1995).
- Lemoine, M., Ferlet, R., Vidal-Madjar, A., Emerich, C. & Bertin, P. Interstellar lithium and the ${}^7\text{Li}/{}^6\text{Li}$ ratio toward ρ Oph. *Astron. Astrophys.* **269**, 469–476 (1993).
- Lemoine, M., Ferlet, R. & Vidal-Madjar, A. The interstellar ${}^7\text{Li}/{}^6\text{Li}$ ratio. The line of sight to ζ Ophiuchi. *Astron. Astrophys.* **298**, 879–893 (1995).
- Morton, D. C. Atomic data for resonance absorption lines. I. Wavelengths longward of the Lyman limit. *Astrophys. J. Suppl.* **77**, 119–202 (1991).
- Federman, S. R., Lambert, D. L., Cardelli, J. A. & Sheffer, Y. The boron isotope ratio in the interstellar medium. *Nature* **381**, 764–766 (1995).
- Lambert, D. L. *et al.* The ${}^{11}\text{B}/{}^{10}\text{B}$ ratio of local interstellar diffuse clouds. *Astrophys. J.* **494**, 614–622 (1998).
- Meyer, D. M., Hawkins, I. & Wright, E. L. The interstellar ${}^7\text{Li}/{}^6\text{Li}$ isotope ratio toward ζ Ophiuchi and ζ Persei. *Astrophys. J.* **409**, L61–L64 (1993).
- van Dishoeck, E. F. & Black, J. H. Comprehensive models of diffuse interstellar clouds: Physical conditions and molecular abundances. *Astrophys. J. Suppl.* **62**, 109–145 (1986).
- Federman, S. R., Weber, J. & Lambert, D. L. Cosmic ray-induced chemistry toward Perseus OB2. *Astrophys. J.* **463**, 181–190 (1996).
- Crane, P., Lambert, D. L. & Sheffer, Y. A very high resolution survey of interstellar CH and CH^+ . *Astrophys. J. Suppl.* **99**, 107–120 (1995).
- White, R. E. Interstellar lithium: differential depletion in diffuse clouds. *Astrophys. J.* **307**, 777–786 (1986).
- Andersen, J., Gustafsson, B. & Lambert, D. L. The lithium isotope ratio in F and G stars. *Astron. Astrophys.* **136**, 65–73 (1984).
- Parizot, E. & Drury, L. Superbubbles as the source of ${}^6\text{Li}$, Be, and B in the early Galaxy. *Astron. Astrophys.* **349**, 673–684 (1999).
- Snow, T. P. Jr Interstellar molecular abundances toward omicron Persei. *Astrophys. J.* **201**, L21–L24 (1975).
- Sansonet, C. J., Richou, B., Engleman, R. Jr & Radziemski, L. J. Measurements of the resonance lines

- of ${}^6\text{Li}$ and ${}^7\text{Li}$ by Doppler-free frequency-modulation spectroscopy. *Phys. Rev. A* **52**, 2682–2688 (1995).
- Hobbs, L. M., Thorburn, J. A. & Rebull, L. M. Lithium isotope ratios in halo stars. III. *Astrophys. J.* **523**, 797–804 (1999).
- Sofia, U. J., Cardelli, J. A., Guerin, K. P. & Meyer, D. M. Carbon in the diffuse interstellar medium. *Astrophys. J.* **482**, L105–L108 (1997).
- Cardelli, J. A., Meyer, D. M., Jura, M. & Savage, B. D. The abundance of interstellar carbon. *Astrophys. J.* **467**, 334–340 (1996).
- Savage, B. D., Bohlin, R. C., Drake, J. F. & Budich, W. A survey of interstellar molecular hydrogen. *Astrophys. J.* **216**, 291–307 (1977).
- Diplas, A. & Savage, B. D. An IUE survey of interstellar H I Ly α absorption. I. Column densities. *Astrophys. J. Suppl.* **93**, 211–228 (1994).
- Federman, S. R. *et al.* Chemical transitions for interstellar C_2 and CN in cloud envelopes. *Astrophys. J.* **424**, 772–792 (1994).

Acknowledgements

We thank E. Parizot and H. Reeves for fruitful discussions. This work was supported by the National Aeronautics and Space Administration. We made use of the Simbad database, operated at Centre de Données Astronomiques de Strasbourg, Strasbourg, France.

Correspondence and requests for materials should be addressed to S.R.F. (e-mail: sfederm@uoft02.utoledo.edu).

Correlated electron emission in multiphoton double ionization

Th. Weber*, H. Giessen†, M. Weckenbrock*, G. Urbasch†, A. Staudte*, L. Spielberger*, O. Jagutzki*, V. Mergel*, M. Vollmer† & R. Dörner*

* Institut für Kernphysik, Universität Frankfurt, August Euler Strasse 6, D-60486 Frankfurt, Germany

† Fachbereich Physik, Philipps-Universität, Renthof 5, D-35032 Marburg, Germany

Electronic correlations govern the dynamics of many phenomena in nature, such as chemical reactions and solid state effects, including superconductivity. Such correlation effects can be most clearly investigated in processes involving single atoms. In particular, the emission of two electrons from an atom—induced by the impact of a single photon¹, a charged particle² or by a short laser pulse³—has become the standard process for studies of dynamical electron correlations. Atoms and molecules exposed to laser fields that are comparable in intensity to the nuclear fields have extremely high probabilities for double ionization^{4,5}; this has been attributed to electron–electron interaction³. Here we report a strong correlation between the magnitude and the direction of the momentum of two electrons that are emitted from an argon atom, driven by a femtosecond laser pulse (at 38 TW cm^{-2}). Increasing the laser intensity causes the momentum correlation between the electrons to be lost, implying that a transition in the laser–atom coupling mechanism takes place.

The interaction of superstrong laser fields with matter has a wide range of applications, from surgery to ignition of nuclear fusion; however, there are still many open questions concerning the underlying fundamental physical processes of this interaction. An extremely clean way to address them is to place just one single atom in the laser focus and examine its reaction. In many cases, the response of the atom to the strong laser field can be well understood by considering only one of its electrons to be active; however, one important experimental observation withstands explanation in such simple terms: the production of doubly and multiply charged ions is more effective by many orders of magnitude than predicted by an independent-electron model^{4,5}. After this discovery, it was agreed that models were needed that include electron–electron correlation to explain this enhanced multiple ionization rate. Intense efforts to model the two-electron process of double ionization in a laser field have reproduced the main feature of a knee

structure in the double ionization yield as a function of laser peak intensity^{6–11} and, moreover, yielded quantitative agreement with the experiment in some cases^{12,13}.

The physical mechanism behind this nonsequential process is, however, still debatable. The most widely accepted model is a re-scattering mechanism^{6,14} in which the first electron is accelerated by the laser field and later driven back by the field towards its parent ion. The second electron is then liberated in an electron–ion impact ionization (for example, see refs 6, 8–10 and 15 for arguments supporting this model and refs 7 and 16 for those opposing it). Alternatively, Fittinghoff *et al.*⁴ suggested that the second electron is shaken off by the same mechanism as in double ionization by a single high-energy photon (for example, from synchrotron radiation). Recently, Becker and Faisal^{12,17} suggested a model of an energy-sharing mechanism using Coulomb correlation represented by a single Feynman diagram. Strong experimental evidence supporting the re-scattering/energy-sharing mechanism is provided by the suppression of double ionization in elliptically polarized laser fields¹⁵.

Here we report the direct experimental observation of strong electron–electron correlation in the final state. We show that the momentum of one electron strongly depends on the momentum of the other electron. Such coincident electron momentum distributions serve as a detailed testing ground for the opposing theoretical approaches. But they also yield direct and intuitive insight into the dynamics of the laser–atom interaction and shed light on the mechanism leading to double ionization. They answer the question of whether the two electrons are pulled by the laser field to the same side and whether the electron–electron repulsion pushes them to opposite sites. The latter is observed with synchrotron radiation at high photon energies, where the energy of just one photon is sufficient to liberate both electrons¹.

A first step towards differential information on the correlated multiple ionization in strong fields has been gained by observing the momentum distribution of the doubly charged ions for helium¹⁸, argon¹⁹ and neon targets²⁰. All three studies found large ion momenta which do not peak at zero, highlighting the correlated nature of the process. In contrast, single ionization, and thus any uncorrelated process¹⁹, gives maximum ion yield at zero momentum.

We have used the well-established technique of cold target recoil ion momentum spectroscopy (COLTRIMS; see refs 21 and 22 for

reviews) for the experiment. The linearly polarized light from a Ti-sapphire laser (800 nm, 220-fs pulse width) was focused by a 5-cm focal length lens onto a pre-cooled supersonic argon gas jet.

The ions and electrons were guided by an electric field to opposite sides towards two position-sensitive channel plate detectors with delay-line readout. Their momenta were determined from the flight times and the positions of impact. For more details of the experimental setup and data analysis see ref. 23. The polarization axis of the light was parallel to the electric field and perpendicular to the direction of the gas jet. The peak laser intensity was determined by measurements of ion yield to an estimated accuracy of 15%.

The local argon pressure in the gas jet was adjusted such that the average ion count rate was about 0.04–0.08 per shot. The simultaneously measured coincidence between Ar^+ ions and electrons allowed an online monitoring of the fraction of false coincidences, in which the detected ion and electron are produced in the same laser shot but do not come from the same atom. Such events (about 30% of the total counts) are localized outside the diagonal line in Fig. 1. The contribution of false coincidences has been subtracted from the Ar^{2+} -electron coincidence data; however, this background correction changes only the details, and not the principal findings of our results.

The localization of the correlated events on the diagonal line shows that the momenta of the Ar^+ ion and electron are equal and have opposite signs for each single event. This proves that the laser pulse is short enough to prevent the electron from leaving the focus during the pulse. The momentum transfer from the optical field to the electron and Ar^+ ion is therefore equal and opposite, at every individual time interval and thus for the whole pulse duration. The laser does not pass on significant net momentum to the electron–ion system. The photon momentum is negligible on the scale of interest here, as 1.5 eV photons have a momentum of only 0.0004 a.u. (atomic units: 1 a.u. of momentum is equivalent to $1.995 \times 10^{-24} \text{ kg m s}^{-1}$). This is essential in this experiment, because it ensures that for double ionization the ion momentum $\mathbf{k}_{\text{ion}}$ is given by the sum of the two electron momentum vectors $\mathbf{k}_{\text{e1,2}}$, $\mathbf{k}_{\text{ion}} = -(\mathbf{k}_{\text{e1}} + \mathbf{k}_{\text{e2}})$. We therefore measured $\mathbf{k}_{\text{ion}}$ and the momentum of one of the electrons and could then deduct the momentum of the second electron from momentum conservation. The horizontal line structure that is visible in Fig. 1 results from above-threshold ionization peaks (for example, ref. 24). These peaks are resolved for slow electrons, whereas they are neither resolved for

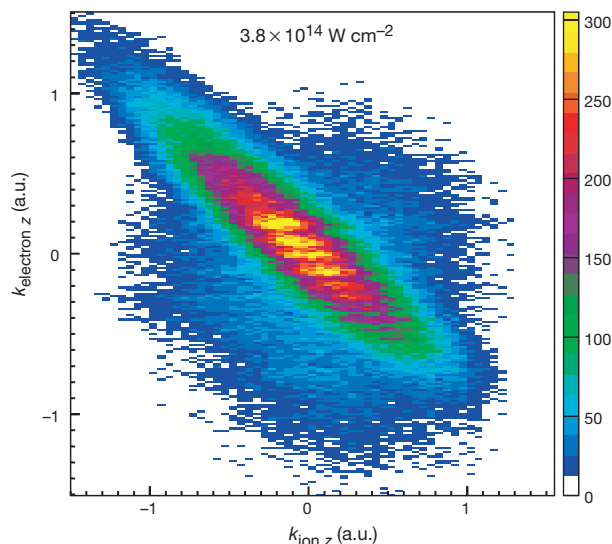


Figure 1 Momentum correlation between the Ar^+ ion and electrons created in the focus of a 220-fs, 800-nm laser pulse at peak intensities of $3.8 \times 10^{14} \text{ W cm}^{-2}$. The horizontal axis shows the momentum component of the Ar^+ ion along the polarization of the laser

field; the vertical axis shows the corresponding momentum component of the detected electron.

the ion momenta nor for faster electrons, because we used a time-of-flight technique. The overall resolution (the width of the diagonal of momentum conservation) is about 0.4 a.u. (full-width at half-maximum).

The main result of our work is shown in Fig. 2. The vertical axis shows the momentum component of the measured electron in the direction of polarization (k_{ez1}) and the horizontal axis the same momentum component of the second electron. The data are integrated over the transverse momentum of the first and second electron. The momentum of the second electron is deduced from the measured momenta of the first electron and the recoiling ion by momentum conservation.

At $3.8 \times 10^{14} \text{ W cm}^{-2}$, right at the 'knee' of the double ionization rate (see refs 19 and 25), a strong momentum correlation between the two electrons is found. There is a clear maximum for both electrons being emitted with the same momentum component of about 1 a.u., whereas emission to opposite half planes is strongly suppressed. At an intensity of $15 \times 10^{14} \text{ W cm}^{-2}$ (Fig. 2b), in the regime where the ratio of double to single ionization steeply rises with laser peak power, the correlation between the electrons is completely lost. Because only one component of the electron momenta is shown, Fig. 2a does not imply that the electrons are really emitted in parallel. More likely, there is an angle between the electrons owing to the repulsion.

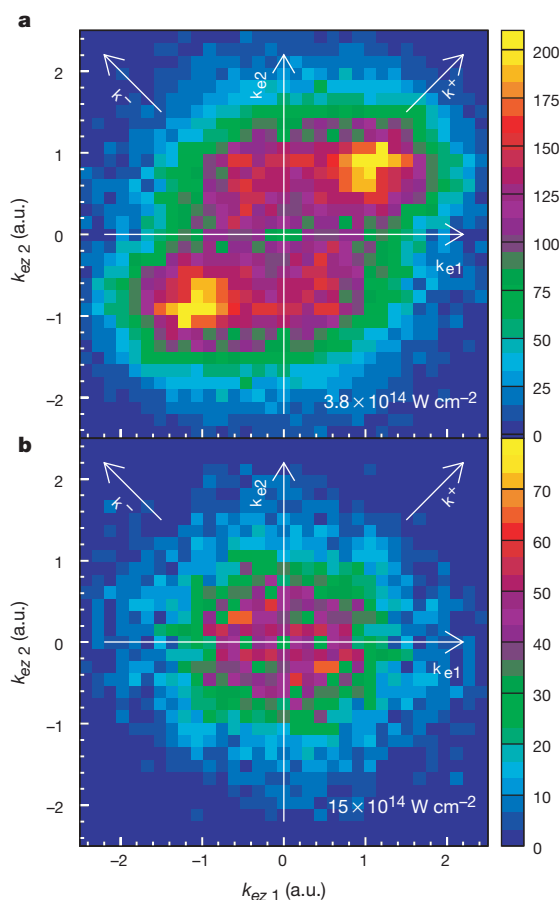


Figure 2 Momentum correlation between the two emitted electrons when an Ar^{2+} ion is produced in the focus of a 220-fs, 800-nm laser pulse at peak intensities of $3.8 \times 10^{14} \text{ W cm}^{-2}$ (a) and $15 \times 10^{14} \text{ W cm}^{-2}$ (b). The horizontal axis shows the momentum component of one electron along the polarization of the laser field; the vertical axis shows the same momentum component of the corresponding second electron. The same sign of the momenta for both electrons means emission to the same half sphere. The data are integrated over the momentum components in the direction perpendicular to the polarization. The colour coding shows the differential rate in arbitrary units.

The difference between Fig. 2a and b, which is accompanied by a sixfold increase in the ratio of double to single ionization^{19,25}, is a direct experimental proof for the change of mechanisms for double ionization as a function of laser power. In the region of the knee in the double ionization rate, the two-electron process is mediated by electron–electron correlation, whereas at high laser powers, double ionization proceeds by sequential emission of the two electrons. As there may be a random number of optical halfcycles between the emission of the two electrons in this case, no directional correlation between the electrons can be expected.

An instructive alternative perspective on Fig. 2 is obtained by rotating the distribution by 45° . The diagonal coordinate frame as indicated in the figure shows the sum and difference momenta $k^+ = k_1 + k_2$ and $k^- = k_1 - k_2$ of the individual electron momenta $k_{1,2}$. Owing to momentum conservation, k^+ is equal and opposite to the Ar^{2+} recoil-ion momentum (compare refs 18–20). Figure 2a shows that the width of the k^- distribution is narrower than the width in the k^+ direction. The coordinates k^+ and k^- are helpful to illustrate the relative importance of the two counteracting effects of electron–electron repulsion and acceleration of particles by the optical field. Both influence the observed final state momenta in different ways. Electron repulsion (and two-body electron–electron scattering) does not change k^+ but contributes to the momentum k^- . On the other hand, once both electrons are set free, the momentum transfer received from the field is identical. Therefore, this part of the acceleration does not change k^- but adds to k^+ . The observed wide k^+ and narrow k^- distributions thus indicate that the joint acceleration of the electrons in the laser field clearly dominates over the influence of electron repulsion. Both electrons are driven by the laser field to the same side. This experimental finding is in agreement with theoretical results calculated using the model of the energy-sharing mechanism²⁶.

How does this finding relate to the two correlation mechanisms of re-scattering and shake-off? The pure case for the latter is found in single-photon double ionization, where the final state momentum distribution is predominantly determined by the long-range Coulomb interaction in the final state. It forces the electrons almost completely to opposite half spheres (see ref. 1 for a review). At small energies, which dominate in our case, k^+ is much smaller than k^- (see ref. 27). If a shake-off process were to take place in a strong laser field, it would set both electrons free in a time interval that would be very short compared with the optical cycle. Furthermore, shake-off would have its highest probability at the maximum intensity of the electric field, which is where single ionization also maximizes. After being set free the electrons perform a quiver motion in the field. The remaining net momentum transfer at the end of the laser pulse depends on the phase at which the electrons are set free. Ionization at the maximum of the field corresponds to zero net momentum transfer from the field. Therefore, the field is not expected to have a significant role in the final state electron momenta resulting from a shake-off. Such electrons should be found either close to the origin in Fig. 2 or in the region of opposite momenta due to electron repulsion. Thus Fig. 2a is in clear contradiction to the expectations of a shake-off model.

In the re-scattering or 'energy-sharing model', there is one particular process that produces electrons with similar momentum. If the energy of the primary electron at its re-encounter with the ion is equal to the energy needed to excite the second electron in the ion, the excitation process would stop the primary electron. The excited electron would then be set free with very little excess energy by the laser. Both electrons will finally have similar momentum which is only determined by the phase of the field at the re-encounter. The momentum of about 1 a.u. corresponds to a phase at re-scattering of 35° and a return energy of the primary electron of about 18 eV. This is sufficient for an excitation, as a broad band of many excited states of the Ar^+ ion starts at about 16.5 eV (the ionization energy is 27 eV). We conclude that the re-scattering model provides a

plausible scheme to produce the observed momentum correlation of the electrons.

Our measured momentum correlation shows that it is not the electron correlation in the initial ground state of the atom that leads to the enhanced double ionization in strong fields, but rather a re-scattering which is a result of the driving field. The magnitude of the momenta of the electrons in the direction of the laser field provides information on the phase of the field at the instant of recollision. The width of the distribution in the k^- direction is sensitive to the details of the re-scattering. Thus we expect that the momentum in the k^- direction and the momentum correlation in the directions perpendicular to the polarization will give further insight into the details of the correlation mechanism in future studies.

Finally, we note that the observed emission of the electrons to the same side confirms a prediction of Taylor *et al.*²⁸. They have solved the time-dependent Schrödinger equation for two electrons in an optical field in three dimensions and found that most of the double ionization probability flux emerges to the same side. Similar conclusions have been drawn from one-dimensional model calculations^{29,30}.

The findings of this work provide a benchmark for theoretical efforts to calculate one of the most simple correlated two-electron processes in nature. Our method, using many-particle momentum space imaging for strong field physics, opens up the road for a variety of applications. The study of electron correlation in strong fields is now possible in molecules, clusters and solids (such as superconductors). □

Received 13 January; accepted 17 March 2000.

- Briggs, J. S. & Schmidt, V. Differential cross sections for photo-double-ionization of the helium atom. *J. Phys. B* **33**, R1–R48 (2000).
- Moshammer, R. *et al.* Double ionization of helium and neon for fast heavy-ion impact: Correlated motion of electrons from bound to continuum states. *Phys. Rev. Lett.* **77**, 1242–1245 (1996).
- Lambropoulos, P., Maragakis, P. & Zhang, J. Two-electron atoms in strong fields. *Phys. Rep.* **305**, 203–293 (1998).
- Fittinghoff, D. N., Bolton, P. R., Chang, B. & Kulander, K. C. Observation of nonsequential double ionization of helium with optical tunneling. *Phys. Rev. Lett.* **69**, 2642–2645 (1992).
- Walker, B. *et al.* Precision measurement of strong field double ionization of helium. *Phys. Rev. Lett.* **73**, 1227–1231 (1994).
- Corkum, P. B. Plasma perspective on strong field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Kulander, K. C., Cooper, J. & Schafer, K. J. Laser-assisted inelastic rescattering during above threshold ionization. *Phys. Rev. A* **51**, 561–568 (1995).
- Liu, W. C., Eberly, J. H., Haan, S. L. & Grobe, R. Correlation effects in two-electron model atoms in intense laser fields. *Phys. Rev. Lett.* **85**, 520–523 (1999).
- Watson, J. B., Sanpera, A., Lappas, D. G., Knight, P. L. & Burnett, K. Nonsequential double ionization of helium. *Phys. Rev. Lett.* **78**, 1884–1887 (1997).
- Lappas, D. G. & van Leeuwen, R. Electron correlation effects in the double ionization of He. *J. Phys. B* **31**, L249–256 (1998).
- LaGattuta, K. J. & Cohen, J. S. Quasiclassical modelling of helium double photoionisation. *J. Phys. B* **31**, 5281–5291 (1998).
- Becker, A. & Faisal, F. H. M. Mechanism of laser-induced double ionization of helium. *J. Phys. B* **29**, L197–202 (1996).
- Becker, A. & Faisal, F. H. M. S-matrix analysis of ionization yields of noble gas atoms at the focus of Ti:sapphire laser pulses. *J. Phys. B* **32**, L335–L343 (1999).
- Schafer, K. J., Yang, B., DiMauro, L. F. & Kulander, K. C. Above threshold ionization beyond the high harmonic cutoff. *Phys. Rev. Lett.* **70**, 1599–1602 (1993).
- Fittinghoff, D. N., Bolton, P. R., Chang, B. & Kulander, K. C. Polarization dependence of tunneling ionization of helium and neon by 120-fs pulses at 614 nm. *Phys. Rev. A* **49**, 2174–2177 (1994).
- Sheehy, B., Lafon, R., Widmer, M., Walker, B. & DiMauro, L. F. Single- and multiple-electron dynamics in the strong-field tunneling limit. *Phys. Rev. A* **58**, 3942–3952 (1998).
- Becker, A. & Faisal, F. H. M. Interplay of electron correlation and intense field dynamics in the double ionization of helium. *Phys. Rev. A* **59**, R1742–R1745 (1999).
- Weber, T. *et al.* Recoil-ion momentum distributions for single and double ionization of helium in strong laser fields. *Phys. Rev. Lett.* **84**, 443–446 (2000).
- Weber, T. *et al.* Sequential and nonsequential contributions to double ionization in strong laser fields. *J. Phys. B* **33**, L128–L133 (2000).
- Moshammer, R. *et al.* Momentum distributions of Ne ions created by an intense ultrashort laser pulse. *Phys. Rev. Lett.* **84**, 447–450 (2000).
- Dörner, R. *et al.* Cold target recoil ion momentum spectroscopy. *Phys. Rep.* **330**, 95–192 (2000).
- Ullrich, J. *et al.* Cold target recoil ion momentum spectroscopy. *J. Phys. B* **30**, 2917–2974 (1997).
- Dörner, R. *et al.* Photo double ionization of He: Fully differential and absolute electronic and ionic momentum distributions. *Phys. Rev. A* **57**, 1074–1090 (1998).
- DiMauro, L. F. & Agostini, P. in *Advances in Atomic and Molecular Physics* 79–120 (Academic, New York, 1995).
- Laroche, S., Talebpour, A. & Chin, S. L. Non-sequential multiple ionization of rare gas atoms in a Ti:Sapphire laser field. *J. Phys. B* **31**, 1201–1214 (1998).

- Becker, A. & Faisal, F. H. M. Interpretation of momentum distribution of recoil ions from laser induced non-sequential double ionization. *Phys. Rev. Lett.* **84**, 3546–3549 (2000).
- Dörner, R. *et al.* Fully differential cross sections for double photoionization of He near threshold measured by recoil ion momentum spectroscopy. *Phys. Rev. Lett.* **77**, 1024–1027 (1996).
- Taylor, K. T., Parker, J. S., Dundas, D., Smyth, E. & Vitirito, S. Laser driven helium in full-dimensionality. *Laser Phys.* **9**, 98–116 (1999).
- Lein, M., Gross, E. K. U. & Engel, V. On the mechanism of strong-field double photoionisation in the helium atom. *J. Phys. B* **33**, 433–442 (2000).
- Dörner, M. Double ionization in a one-cycle laser pulse. *Optics Express* **6**, 111–116 (2000).

Acknowledgements

We thank H. Schmidt-Böcking for enthusiastic support, and R. Moshammer and J. Ullrich for helpful discussions. Our analysis of the influence of the laser field on the final state momenta emerged after fruitful discussion with A. Becker and F. H. M. Faisal. We are grateful to W. W. Rühle for continuous support and thank Roentdek GmbH for providing the position sensitive detectors. This work is supported by DFG, BMBF, GSI and DAAD. R.D. is supported by the Heisenberg Programme of the DFG. The Marburg group thanks the Land Hessen and the DFG for support through the SFB383 and their Graduiertenkolleg 'Optoelektronik mesoskopischer Halbleiter'.

Correspondence and requests for materials should be addressed to R.D. (e-mail: doerner@hsb.uni-frankfurt.de).

Improving the performance of doped π -conjugated polymers for use in organic light-emitting diodes

Markus Gross*, David C. Müller*, Heinz-Georg Nothofer†, Ulrich Scherf†, Dieter Neher‡, Christoph Bräuchle* & Klaus Meerholz*

* Institut für Physikalische Chemie, Ludwig-Maximilians-Universität München, Butenandtstrasse 11, 81377 München, Germany

† Max-Planck Institute für Polymerforschung, Institut Ackermannweg 10, 55037 Mainz, Germany

‡ Institut für Experimentalphysik, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Organic light-emitting diodes (OLEDs) represent a promising technology for large, flexible, lightweight, flat-panel displays^{1–3}. Such devices consist of one or several semiconducting organic layer(s) sandwiched between two electrodes. When an electric field is applied, electrons are injected by the cathode into the lowest unoccupied molecular orbital of the adjacent molecules (simultaneously, holes are injected by the anode into the highest occupied molecular orbital). The two types of carriers migrate towards each other and a fraction of them recombine to form excitons, some of which decay radiatively to the ground state by spontaneous emission. Doped π -conjugated polymer layers improve the injection of holes in OLED devices^{4–9}; this is thought to result from the more favourable work function of these injection layers compared with the more commonly used layer material (indium tin oxide). Here we demonstrate that by increasing the doping level of such polymers, the barrier to hole injection can be continuously reduced. The use of combinatorial devices allows us to quickly screen for the optimum doping level. We apply this concept in OLED devices with hole-limited electroluminescence (such as polyfluorene-based systems^{10–12}), finding that it is possible to significantly reduce the operating voltage while improving the light output and efficiency.

OLED devices are typically fabricated using one transparent electrode. The material most commonly employed for this purpose is indium tin oxide (ITO), which serves as the anode. The electronic work function ϕ_w of ITO is generally smaller than the highest

occupied molecular orbital (HOMO) level of most organic semiconductors, resulting in a barrier ϕ_h for hole injection. The work function of ITO has been shown to depend strongly on the manufacturing process and any pretreatment^{13,14}, which makes the reproducibility of device fabrication rather difficult. Heeger *et al.*⁵ have shown that coating the ITO with “doped” (that is, oxidized) polyaniline reduced ϕ_h and, consequently, lowered the onset voltage and operating voltage of the devices. More recently, polypyrrole (ref. 6) and most importantly poly(3,4-ethylenedioxythiophene) (PEDOT) have been used as the injection layer for holes^{7–9}. We note that counter ions are needed for compensation of the charges of the

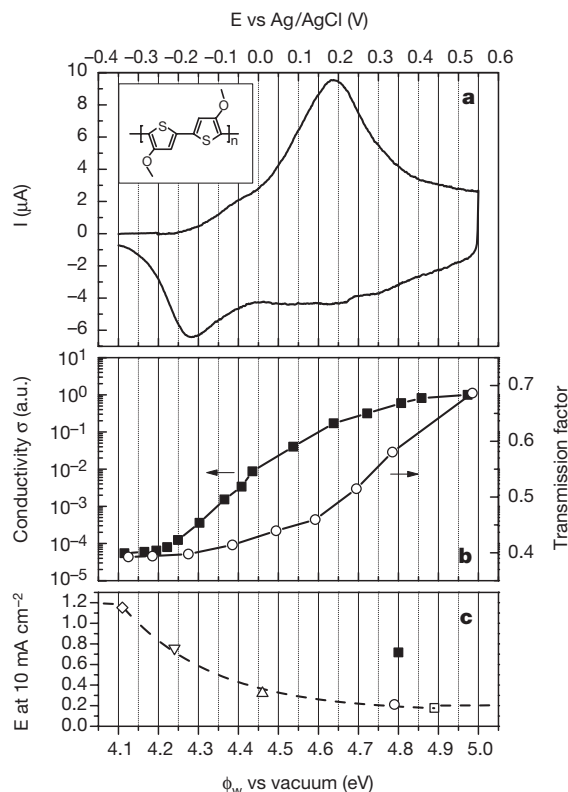


Figure 1 Electrochemical and electro-optical properties of PDBT and its performance as the hole injection contact of hole-only OLED devices. **a**, Cyclic voltammogram of a PDBT film deposited on ITO, which served as the working electrode, in the section of interest for this study. The measurement was performed in a 0.1 M tetrabutylammonium hexafluorophosphate/acetonitrile solution. The upper x axis shows the work function ϕ_w versus the vacuum level. The offset between the two axes is 4.45 eV (ref. 16). Note that variations of ϕ_w cannot be achieved with PDBT below 4.1 eV due to the absence of redox couples (no current in the cyclic voltammogram). Inset, chemical structure of PDBT. **b**, Conductivity σ (filled squares, left axis) and transmission factor for PFO emission (open circles, right axis) of 150-nm PDBT films as a function of the adjusted potential E_{eq} /work function ϕ_w . The relative conductivity σ was obtained by investigation of a PDBT film deposited on a pair of interdigitated electrodes. For each particular data point, the film was adjusted to a certain redox potential in an acetonitrile/NaPSS solution, dried, and then the impedance (equivalent to resistance for low frequency) was determined by measuring the voltage drop on a calibrated resistor in series with the film by use of a lock-in amplifier. The total applied a.c. voltage was 10 mV and the frequency was 4 Hz. The data were normalized to the value obtained for $E_{eq} = +0.53$ V. The transmission factor was obtained by convolution of each individual absorption spectrum for a given E_{eq} with the EL spectrum of PFO (inset Fig. 3b), integration, and finally normalization with the integrated EL spectrum through ITO. The absorption spectra of PDBT were measured *in situ* in a transmission electrochemical cell as a function of E_{eq} . **c**, Electric field necessary to pass a current of 10 mA cm⁻² through the ‘hole-only’ devices with the general structure anode/hole-transport layer/Al determined from the data in Fig. 2. This field is plotted against the adjusted potential E_{eq} /work function ϕ_w for the devices with different PDBT injection layers (open symbols) and the device using a PEDOT anode (filled square). E_{eq} was determined before spin-casting the hole-transport layer. The dashed line is a guide to the eye.

oxidized π -conjugated polymer; such counter ions might also have a considerable influence. In the case of commercially available PEDOT, the polyanion poly(styrene sulphonate) (PSS) is commonly used.

All studies using π -conjugated polymeric anodes for OLED applications have so far neither reported nor taken into account the exact doping level of the polymer, though it has long been known that the latter is correlated with the electrochemical equilibrium potential of the polymer, E_{eq} (ref. 15). Generally, a simple offset of 4.45 eV is assumed to exist between electrochemical potentials and distinct energy levels such as the work function (ϕ_w) (refs 10, 16). Ultraviolet Photon Spectroscopy (UPS) measurements performed on two differently synthesized PEDOT samples yielded ϕ_w values of ~ 4.0 eV for ‘neutral’ and 4.4 eV for ‘doped’ material, respectively¹⁷. More recently, $\phi_w = 5.2$ –5.3 eV was determined by electroabsorption and the Kelvin-probe method¹⁸. These large variations may not only be caused by the different methods, but probably also stem from differences in the film preparation leading to different degrees of oxidation (work function) of the PEDOT. Because ϕ_w directly influences the barrier ϕ_h , strong variations in the OLED performance are expected.

These considerations clearly indicate that variations in the degree of oxidation might be one of the possible sources for the irreproducible performance of OLED devices using polymeric hole-injection contacts. On the other hand, they also suggest that it should be possible to actively influence ϕ_w of such anodes by adjusting the doping level of the polymer. The particular advantage of π -conjugated polymers for this purpose is that they are generally redox-active over a broad potential range¹⁵. Within this range, a continuous distribution of redox states is observed as is clearly evident from the cyclic voltammogram in Fig. 1a. Thus, the work function ϕ_w should be adjustable in infinitesimally small steps within the redox-active range by either electrochemical or chemical means.

Here we introduce poly(4,4'-dimethoxy-bithiophene) (PDBT, inset Fig 1a) as a ‘new’ hole-injection contact material for OLEDs, because it offers several advantages over other π -conjugated polymers: (1) it is the most stable conjugated polymer known to date,

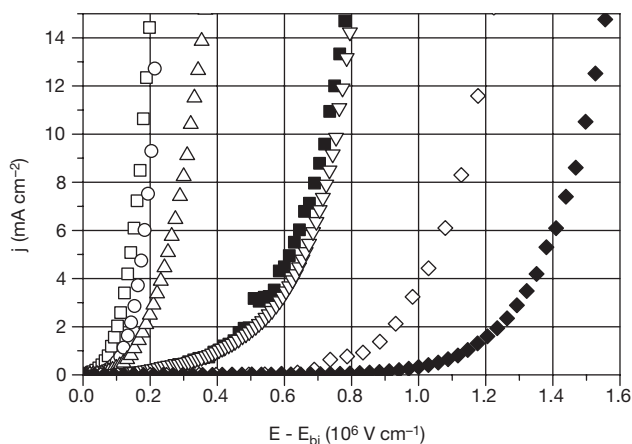


Figure 2 I–V characteristics of hole-only devices with differently doped PDBT injection contacts. The figure shows the field dependence of the electric current density j for various devices with the general structure anode/hole-transport layer/Al. PDBT devices adjusted to different electrochemical potentials E_{eq} /work functions ϕ_w (open symbols): $E_{eq} = +0.44$ V (squares), $+0.34$ V (circles), $+0.01$ V (up triangles), -0.21 V (down triangles), and -0.34 V (diamonds). PEDOT device with $E_{eq} = 0.36$ V (filled squares) and plain ITO device (filled diamonds). All E_{eq} values are given versus the reference electrode Ag/AgCl. The polymeric anodes were 40–50 nm thick, the hole-transport layer 200-nm thick. In all cases, the voltage actually applied (or electric field) was corrected for the built-in potential V_{bi} resulting from the different work functions of the two contact electrodes^{18,21}.

that is, it is in principle fully oxidizable (one charge per thiophene unit) without chemical degradation¹⁹; (2) it offers an extremely wide range of tunability for the electrochemical equilibrium potential E_{eq} of between -0.4 V and $+2.5$ V versus Ag/AgCl (ref. 19) (4.05 eV $< \phi_w < 6.95$ eV; note that Fig. 1a only shows the section of interest for this study). Additionally, PDBT is almost transparent to visible light in the doped state, and exhibits electrical conductivity $\sigma > 100$ S cm⁻¹.

We employed three kinds of hole-injecting anodes in our devices: (1) ITO (20 Ω per square), (2) PEDOT films on ITO-coated substrates spin-coated from an aqueous suspension ('Baytron P', Bayer Corporation), and (3) PDBT films oxidized at different redox potentials. The latter were deposited onto ITO-coated substrates by electropolymerization of the 'monomer', 4,4'-dimethoxybithiophene, at 1.2 V in a saturated water/acetonitrile 1:1 solution containing sodium poly(styrene sulphonate) (NaPSS, Aldrich; molecular mass, 70,000 g mol⁻¹) as the supporting electrolyte. After preparation, the degree of oxidation of the PDBT films was adjusted electrochemically to different values in a monomer-free acetonitrile/NaPSS solution. Thereafter, the films were carefully rinsed with acetonitrile and water. Finally, all substrates were dried at 80 °C in vacuum. The equilibrium potential E_{eq} was determined after this entire procedure (accuracy ± 0.05 V). For device fabrication, the organic layers were deposited by spin-casting the appropriate solutions onto the as-prepared substrates and then dried in vacuum. We note that PEDOT and PDBT are both insoluble after

film formation. Finally, a top electrode was evaporated, yielding diodes with an active area of 9.1 mm². All devices were characterized in nitrogen atmosphere.

Previous studies using π -conjugated polymers as the hole-injecting anode were restricted to the doped (that is, conductive) state of those polymers, and it was demonstrated that the OLED performance is independent of the thickness of the conjugated polymer⁴⁻⁹. Therefore, it was considered as part of the injection contact. However, it is well-known that the conductivity σ of conjugated polymers generally varies for different doping levels¹⁵, which has to be taken into account, in particular for the state of low conductivity. For PDBT, σ decreases over four orders of magnitude upon discharging (Fig. 1b), but despite this we did not observe a strong influence of the uncharged PDBT layer thickness on the I - V curves. Nevertheless, to keep this ambiguity as small as possible, the PDBT layer was generally less than one-third the thickness of the following semiconducting layer(s).

First, we studied the influence of the anode material on the I - V characteristics in single-layer 'hole only' devices. They consisted of a hole-transport layer containing 50 wt% of N,N' -diphenyl- N,N' -(3-methylphenyl)-1,1'-biphenyl-4,4'-diamine (TPD; Syntec GmbH) dispersed in polycarbonate (Aldrich; molecular mass, 64,000 g mol⁻¹). Aluminum served as the cathode. None of the devices showed electroluminescence (EL) due to the negligibly small number of electrons. As expected the ITO device showed a higher current onset field than all devices using polymeric anodes (Fig. 2), a fact which is typically attributed to the more favourable work function⁴⁻⁹. Within the series of devices using a PDBT anode, the onset field for hole injection decreases systematically with increasing oxidation potential E_{eq} . Any intermediate performance level not specifically shown can be obtained by properly adjusting E_{eq} . As the hole-transport layer is identical in all cases, these observations must be correlated with a decreased barrier ϕ_h as a result of increased work function ϕ_w of the PDBT. They constitute an indirect, but unambiguous, proof for the feasibility of our concept of influencing the work function by adjusting the redox state of a conjugated polymer.

Due to the lack of suitable analytical models describing charge-injection into amorphous organic semiconductors, an exact and quantitative evaluation of the data is at present not possible. Instead, we plotted the field necessary to pass a certain current density through the devices (for example, 10 mA cm⁻²) as a function

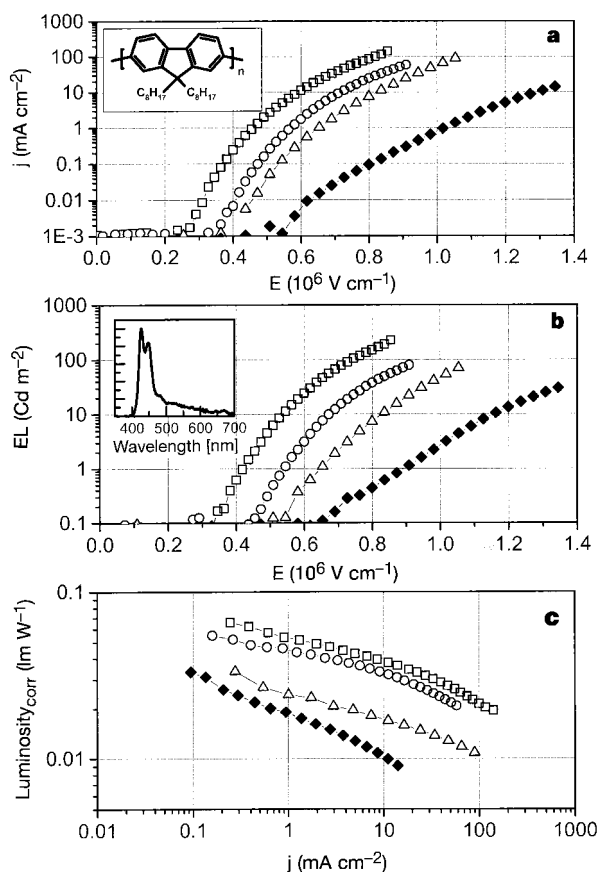


Figure 3 Performance of various OLED devices with the general structure anode/PFO/Ca. Results are shown for PDBT devices adjusted to different electrochemical potentials E_{eq} /work functions ϕ_w (open symbols): $E_{eq} = +0.30$ V (squares), -0.05 V (circles) and -0.25 V (up triangles); ITO reference device (filled diamonds). All E_{eq} values are given versus the reference electrode Ag/AgCl. The polymeric anodes were 150-nm thick, and the hole-transport layer was 550-nm thick. **a**, Field dependence of the electric current density j . Inset, chemical structure of PFO. **b**, Field dependence of the light output (EL). Inset, emission spectrum of PFO. **c**, Current dependence of the luminosity.

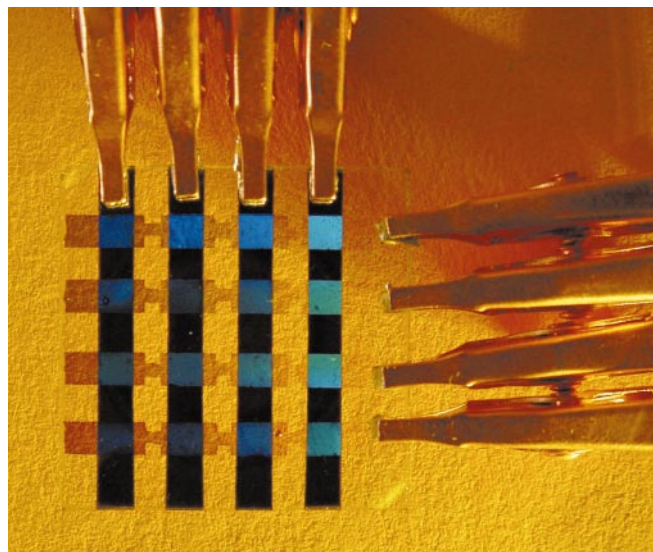


Figure 4 Two-dimensional combinatorial OLED array. The PDBT thickness increases top to bottom; the electrochemical equilibrium potential increases left to right. See text for details.

of E_{eq} (ϕ_w ; see Fig. 1c). It can be seen that this field (E at 10 mA cm^{-2}) decreases dramatically with increasing E_{eq} , and levels off before reaching the HOMO level of TPD (5.07 eV, determined electrochemically in solution). This indicates that even though there should still be a considerable energy barrier between the HOMO of TPD and the work function of the highly-doped PDBT, an ohmic contact has almost been achieved. This is in agreement with a recent theoretical study²⁰ predicting ohmic contacts for energy barriers smaller than 0.3 eV. Charge injection in the PEDOT reference device ($E_{eq}(\text{PEDOT}) = +0.36 \text{ V}$) was consistently more difficult than in the PDBT device displaying a similar E_{eq} of the anode layer. These results indicate that the work function is not the only factor influencing carrier injection and performance of OLED devices. The discrepancy might result from the different preparation methods of the PEDOT and PDBT films (spin-coating of the suspension versus electrochemical deposition), or might be due to the different substituents at the thiophene ring (ethylenedioxy ring versus methoxy group). Note that the dopant counter ions were identical in both cases (PSS).

Because the onset field decreases, and density of holes at a given field increases, with increasing doping level (E_{eq} , ϕ_w) of the PDBT, this should have a very marked effect on the performance of OLED devices. We distinguish two cases; (1) devices with an excess number of holes in the emission layer and (2) hole-deficient devices. A typical example of the more-common first case are single-layer poly(2-methoxy, 5-(2'-ethyl-hexyloxy)-1,4-phenylenevinylene) (MEH-PPV) devices. Here, the increase of the doping level of the PDBT anode led to an even further increased hole current at a given field, while the brightness remained basically unchanged. Overall, this yielded a reduction of the efficiency (M.G., D.C.M. and K.M., unpublished results). To study the opposite case (2), we selected 9,9-bis(n-octyl)fluorene (PFO) as the emitting layer, and again simple one-layer devices were fabricated. PFO has the HOMO at 5.8 eV, giving rise to a large barrier for hole injection from ITO^{10,11}. Calcium was evaporated as the electron-injecting cathode.

All devices showed blue electroluminescence with a maximum wavelength of 436 nm, similar to results reported in ref. 10 (inset Fig. 3). As expected, the onsets of the current and the light output both decrease monotonously with increasing E_{eq} (ϕ_w) of the PDBT anode (Fig. 3). For a given electric field, the EL increases by more than one order of magnitude in the devices using charged PDBT compared with those using non-charged PDBT and even more than two orders of magnitude compared with the ITO reference device despite absorption losses in the PDBT layer (Fig. 1b). Within the series of PDBT devices, the best external light output of 250 Cd m^{-2} was achieved with the highly-doped PDBT device with an external efficiency of $\eta_{ext} = 0.18 \text{ Cd A}^{-1}$. Taking into account the absorption of the PDBT (Fig. 1b), this corresponds internally to $\eta_{corr} \approx 0.33 \text{ Cd A}^{-1}$, which is comparable to the values for a bilayer device using a "TPD-polymer" as hole-transporting layer¹⁰.

In Fig. 3c, the luminosity (as a measure for the energy conversion efficiency of the OLED) is plotted as a function of current. For better comparison, the EL data have been corrected for variations of the PDBT transmission (see Fig. 1b), which generally accompany the change of the redox state in π -conjugated polymers¹⁵. The luminosity of the PDBT devices is dependent on E_{eq} , and is a factor ~ 2.5 times higher in the highly-doped PDBT device than in the slightly-doped case and ~ 4 times higher than for the ITO reference for the same current density (for example, 10 mA cm^{-2}). This shows that the enhanced hole injection not only reduces the onset voltage, but also influences the recombination efficiency positively due to the more balanced number density of the two carrier types inside the luminescent layer. On the other hand, for a given electric field (such as $E = 0.8 \text{ MV cm}^{-1}$) the luminosity exhibits an intermediate optimum value. We attribute the decreased luminosity for higher ϕ_w to enhanced heating phenomena.

We have shown that 'hole-only' and OLED devices using highly-

doped PDBT anodes showed strongly reduced onset fields compared to those using non-doped PDBT, ITO or commercially available PEDOT as the anode. Furthermore, the luminosity in OLEDs based on a polyfluorene derivative, that is, with hole-limited EL, was increased. Similar observations have been obtained with a number of materials with very different HOMO levels, including various composites and emissive conjugated polymers. Our concept thus opens up the possibility of optimizing devices by properly adjusting the work function of the polymeric anode. We point out that this is not restricted to PDBT, but may be used with π -conjugated polymers in general. The alternative way to influence charge injection into a certain organic semiconductor is the introduction of additional transport layer(s). Normally, several layers are necessary to cause the barrier to vanish, and ultimately higher voltages are required, reducing the power efficiency.

We have used electrochemistry for screening purposes, even though it is possible to achieve similar results by chemical oxidation/reduction processes. But electrochemistry is more convenient, as it simultaneously offers the possibility of (1) depositing the polymer onto the ITO, (2) adjusting the degree of oxidation in infinitesimally small steps by simply changing the 'oxidation power' of the working electrode, and (3) determination of E_{eq} before deposition of the other organic layers. Furthermore, combinatorial devices can easily be fabricated, greatly reducing the time necessary for device optimization. Each device on one substrate can be adjusted independently to a different E_{eq} by means of electrochemistry. An example of a two-dimensional combinatorial array is shown in Fig. 4. In the first dimension, the redox state is varied, and in the second dimension, the thickness of the polymeric anode material is changed. Alternatively, various injection polymers with identical E_{eq} could be tested, and so on. Once the optimum doping level is known, the conditions for achieving this specific doping level by chemical methods have to be explored to make our approach applicable for preparation of these films on a large scale in an industrial environment.

It should be possible to extend this principle, which is presented here for hole injection only, to electron injection. Our concept also opens the way to various new possibilities for tuning and structuring displays. For example, it could perhaps be used to fabricate monochrome 'grey-level' displays by adjusting to different redox states the specific areas of a substrate covered uniformly with PDBT. Due to the locally varying hole currents, the luminescent layer would then emit light with different brightnesses. □

Received 3 February; accepted 20 April 2000.

1. Friend, R. H. *et al.* Electroluminescence in conjugated polymers. *Nature* **397**, 121–128 (1999).
2. Heeger, A. J. Light emission from semiconducting polymers: light-emitting diodes, light-emitting electrochemical cells, lasers and white light for the future. *Solid State Commun.* **107**, 673–679 (1998).
3. Salbeck, J. Electroluminescence with organic compounds. *Ber. Bunsenges. Phys. Chem.* **100**, 1667–1677 (1996).
4. Gustafsson, G. *et al.* Flexible light-emitting diodes made from soluble conducting polymers. *Nature* **357**, 477–479 (1992).
5. Yang, Y. & Heeger, A. J. Polyaniline as a transparent electrode for light-emitting diodes: Lower operating voltage and higher efficiency. *Appl. Phys. Lett.* **64**, 1245–1247 (1994).
6. Gao, J., Heeger, A. J., Lee, Y. J. & Kim, C. Y. Soluble polypyrrole as the transparent anode in polymer light-emitting diodes. *Synth. Met.* **82**, 211–223 (1996).
7. Carter, S. A., Angelopoulos, M., Karg, S., Brock, P. J. & Scott, J. C. Polymeric anodes for improved polymer light-emitting diode performance. *Appl. Phys. Lett.* **70**, 2067–2069 (1997).
8. Scott, J. C., Carter, S. A., Karg, S. *et al.* Angelopoulos, M. Polymeric anodes for light-emitting diodes. *Synth. Met.* **85**, 1197–1200 (1997).
9. Cao, Y., Yu, G., Zhang, C., Menon, R. & Heeger, A. J. Polymer light-emitting diodes with polyethylene dioxathiophene-polystyrene sulfonate as the transparent anode. *Synth. Met.* **87**, 171–174 (1997).
10. Grice, A. W. *et al.* High brightness and efficient blue light-emitting polymer diodes. *Appl. Phys. Lett.* **73**, 629–631 (1998).
11. Janiez, S. *et al.* Electrochemical determination of the ionization potential of poly(9,9-diocetylfluorene). *Appl. Phys. Lett.* **73**, 2453–2455 (1998).
12. Grell, M. *et al.* Polarized electroluminescence from a liquid crystalline polyfluorene. *Adv. Mater.* **11**, 671–674 (1999).
13. Nuesch, F., Rothberg, L. J., Forsythe, E. W., Quoc Toan, L. & Gao, Y. A. Photoelectron spectroscopy study on the indium tin oxide treatment by acids and bases. *Appl. Phys. Lett.* **74**, 880–882 (1999).
14. Kim, J. S. *et al.* Indium tin oxide treatments for single and double layer polymeric light emitting diodes. The relation between the anode physical, chemical, and morphological properties and the device performance. *J. Appl. Phys.* **84**, 6859–6870 (1998).

15. Skotheim, T. A. *Handbook of Conducting Polymers* (Dekker, New York, 1986).
16. Bard, A. J. & Faulkner, L. A. *Electrochemical Methods—Fundamentals and Applications* (Wiley, New York, 1984).
17. Xing, K. Z., Fahlman, M., Chen, X. W., Inganäs, O. & Salaneck, W. R. The electronic structure of poly(3,4-ethylene-dioxythiophene): studied by XPS and UPS. *Synth. Met.* **89**, 161–165 (1997).
18. Brown, T. M. *et al.* Built-in field electroabsorption spectroscopy of polymer light-emitting diodes incorporating a doped poly(3,4-ethylene dioxythiophene) hole injection layer. *Appl. Phys. Lett.* **75**, 1679–1681 (1999).
19. Dietrich, M. & Heinze, J. Poly(4,4'-Dimethoxybithiophene)—A new conducting polymer with extraordinary redox and optical properties. *Synth. Met.* **41–43**, 503–506 (1991).
20. Malliaras, G. G. & Scott, J. C. Numerical simulations of the electrical characteristics and the efficiencies of single-layer organic light emitting diodes. *J. Appl. Phys.* **85**, 7426–7432 (1999).
21. Malliaras, G. G., Salem, J. R., Brock, P. J. & Scott, C. Electrical characteristics and efficiency of single-layer organic light-emitting diodes. *Phys. Rev. B* **58**, R13411–R13414 (1998).

Acknowledgements

We thank J. Heinze for making some bithiophene available, W. Brütting for discussions, and Bayer AG (Leverkusen, Germany) for the donation of Baytron P.

Correspondence and requests for materials should be addressed to K.M. (e-mail: klaus.meerholz@cup.uni-muenchen.de).

Selection of peptides with semiconductor binding specificity for directed nanocrystal assembly

Sandra R. Whaley*, D. S. English*, Evelyn L. Hu†, Paul F. Barbara*‡ & Angela M. Belcher*‡

* Department of Chemistry and Biochemistry, ‡ The Texas Materials Institute, The University of Texas at Austin, Austin, Texas 78712, USA

† Center for Quantized Electronic Structures, University of California, Santa Barbara, California 93106, USA

In biological systems, organic molecules exert a remarkable level of control over the nucleation and mineral phase of inorganic materials such as calcium carbonate and silica, and over the assembly of crystallites and other nanoscale building blocks into complex structures required for biological function^{1–4}. This ability to direct the assembly of nanoscale components into controlled and sophisticated structures has motivated intense efforts to develop assembly methods that mimic or exploit the recognition capabilities and interactions found in biological systems^{5–10}. Of particular value would be methods that could be applied to materials with interesting electronic or optical properties, but natural evolution has not selected for interactions between biomolecules and such materials. However, peptides with limited selectivity for binding to metal surfaces and metal oxide surfaces have been successfully selected^{10,11}. Here we extend this approach and show that combinatorial phage-display libraries can be used to evolve peptides that bind to a range of semiconductor surfaces with high specificity, depending on the crystallographic orientation and composition of the structurally similar materials we have used. As electronic devices contain structurally related materials in close proximity, such peptides may find use for the controlled placement and assembly of a variety of practically important materials, thus broadening the scope for 'bottom-up' fabrication approaches.

Phage-display libraries, based on a combinatorial library of random peptides—each containing 12 amino acids—fused to the pIII coat protein of M13 coliphage¹², provided us with ~10⁹ different peptides that were reacted with crystalline semiconductor structures. Five copies of the pIII coat protein are located on one end

of the phage particle, accounting for 10–16 nm of the particle. The phage-display approach provided a physical linkage between the peptide–substrate interaction and the DNA that encodes that interaction. The experiments described here utilized five different single-crystal semiconductors: GaAs(100), GaAs(111)A, GaAs(111)B, InP(100) and Si(100). These substrates allowed for systematic evaluation of the peptide–substrate interactions. Protein sequences that successfully bound to the specific crystal were eluted from the surface, amplified by 10⁶, and re-reacted against the substrate under more stringent conditions. This procedure was repeated five times to select the phage with the most specific binding. After the third, fourth and fifth rounds of phage selection, crystal-specific phage were isolated and their DNA sequenced. We identified peptide binding that is selective for the crystal composition (for example, binding to GaAs but not to Si) and crystalline face (for example, binding to (100) GaAs, but not to (111)B GaAs).

Twenty clones selected from GaAs(100) were analysed to determine epitope binding domains to the GaAs surface. The partial peptide sequences of the modified pIII protein are shown in Fig. 1, revealing similar amino-acid sequences among peptides exposed to GaAs. With increasing number of exposures to a GaAs surface, the number of uncharged polar and Lewis-base functional groups increased. Phage clones from third, fourth and fifth round sequencing contained on average 30%, 40% and 44% polar functional groups, respectively, while the fraction of Lewis-base functional groups increased at the same time from 41% to 48% to 55%. The observed increase in Lewis bases, which should constitute only 34% of the functional groups in random 12-mer peptides from our library, suggests that interactions between Lewis bases on the peptides and Lewis-acid sites on the GaAs surface may mediate the selective binding exhibited by these clones.

The expected structure of the modified 12-mers selected from the library would be an extended conformation, which seems likely for

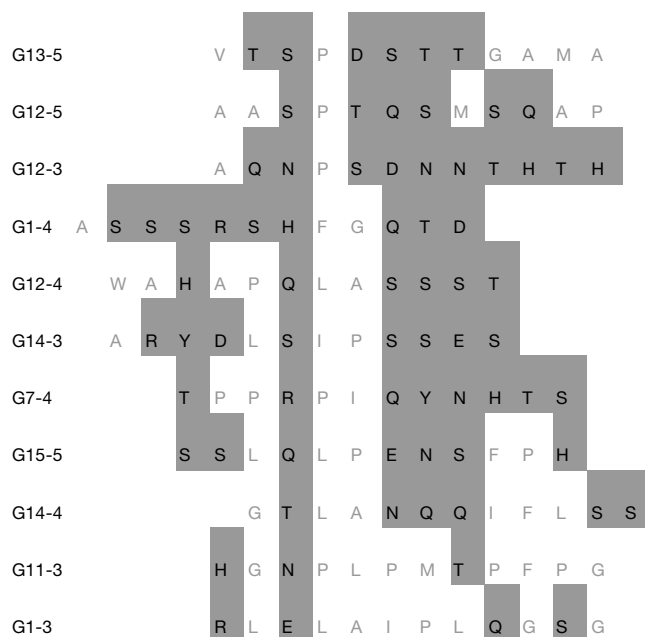


Figure 1 Selected amino-acid sequences of randomized peptide inserts. Here we show partial amino-acid sequences from clones that bound to GaAs(100). These sequences have amino acids with functional groups that can donate electrons to the GaAs surface (shaded areas). Serine (S) and threonine (T) both have hydroxyl side chains that differ by one carbon. Asparagine (N) and glutamine (Q) both have amine-bearing side chains that differ by one carbon. Non-polar amino acids are indicated by grey lettering. The clones were named as follows: G1–3 was the first clone selected from GaAs (100) during the third round of selection.

small peptides, making the peptide much longer than the unit cell (5.65 Å) of GaAs. Therefore, only small binding domains would be necessary for the peptide to recognize a GaAs crystal. These short peptide domains, highlighted in Fig. 1, contain serine- and threonine-rich regions in addition to the presence of amine Lewis bases, such as asparagine and glutamine. To help elucidate the exact binding sequence, we have begun screening our surfaces with shorter libraries, including 7-mer and disulphide constrained 7-mer libraries. By reducing the size and flexibility of the binding domain, fewer peptide–surface interactions are allowed, thereby increasing the strength of interactions between generations of selection.

Phage, tagged with streptavidin-labelled 20-nm colloidal gold particles bound to the phage through a biotinylated antibody to the M13 coat protein, were used for quantitative assessment of specific binding. X-ray photoelectron spectroscopy (XPS) elemental composition determination was performed, monitoring the phage–substrate interaction through the intensity of the gold 4f-electron

signal (Fig. 2a–c). Without the presence of the G1-3 phage, the antibody and the gold streptavidin did not bind to the GaAs(100) substrate. The gold binding was, therefore, specific to the phage and an indicator of the phage binding to the substrate.

Using XPS we showed that the G1-3 clone isolated from GaAs(100) bound specifically to GaAs(100) but not to Si(100) (see Fig. 2a). In complementary fashion the S1 clone, screened against the (100) Si surface, showed poor binding to the (100) GaAs surface. Some GaAs clones also bound the surface of InP (100), another zinc-blende structure. The basis of the selective binding, whether it is chemical, structural or electronic, is still under investigation. In addition, the presence of native oxide on the substrate surface could alter the selectivity of peptide binding.

The preferential binding of the G1-3 clone to GaAs(100), over the (111)A (gallium terminated) or (111)B (arsenic terminated) face of GaAs was shown (Fig. 2b, c). The G1-3 clone surface concentration was greater on the (100) surface, which was used for its selection, than on the gallium-rich (111)A or arsenic-rich (111)B surfaces. These different surfaces are known to exhibit different chemical reactivities, and it is not surprising that there is selectivity demonstrated in the phage binding to the various crystal faces. Although the bulk termination of both 111 surfaces give the same geometric structure, the differences between having Ga or As atoms outermost in the surface bilayer become more apparent when comparing surface reconstructions. The composition of the oxides of the various GaAs surfaces is also expected to be different, and this in turn may affect the nature of the peptide binding.

In Fig. 2c, we plot the intensity of Ga 2p electrons against the binding energy from substrates that were exposed to the G1-3 phage clone. As expected from the results in Fig. 2b, the Ga 2p intensities observed on the GaAs (100), (111)A and (111)B surfaces are inversely proportional to the gold concentrations. This decrease in Ga 2p intensity on surfaces with higher gold concentrations was due to the increase in surface coverage by the phage. XPS is a surface technique with a sampling depth of approximately 30 Å; therefore,

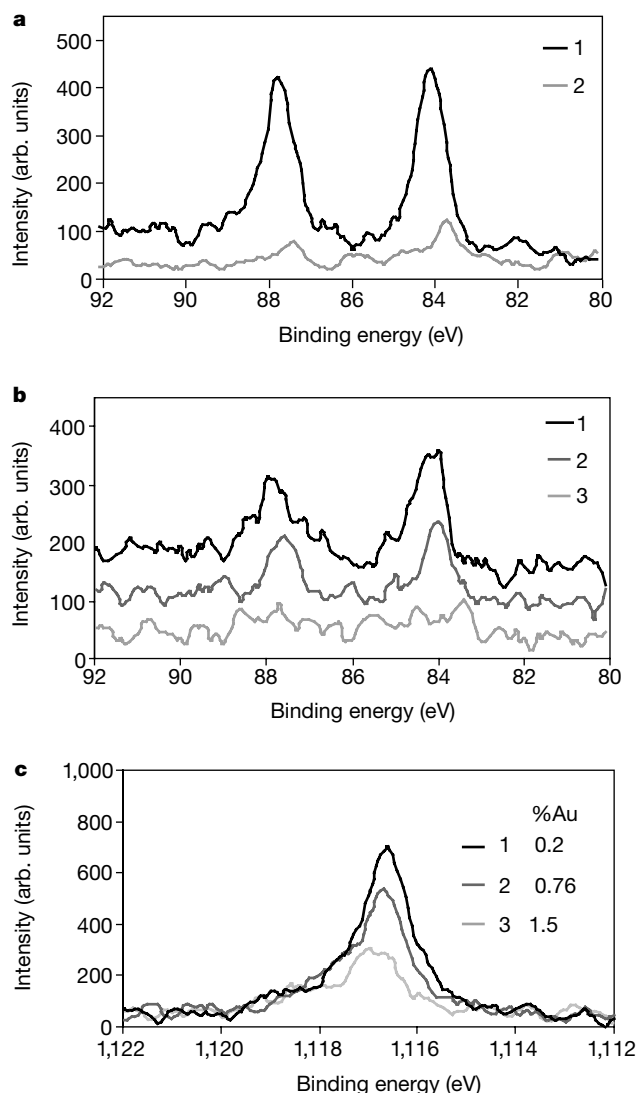


Figure 2 Detection of gold-labelled phage on GaAs using X-ray photoelectron spectroscopy. **a**, Plot of Au 4f binding energy versus intensity for the G1-3 clone on GaAs(100) (curve 1) and Si(100) (curve 2) surfaces. **b**, As **a** but for the following surfaces: GaAs(100) (curve 1), GaAs(111)A (curve 2), and GaAs(111)B (curve 3). **c**, Plot of Ga 2p binding energy versus intensity for the G1-3 clone on the following surfaces: GaAs(111)B (curve 1), GaAs(111)A (curve 2), and GaAs(100) (curve 3).

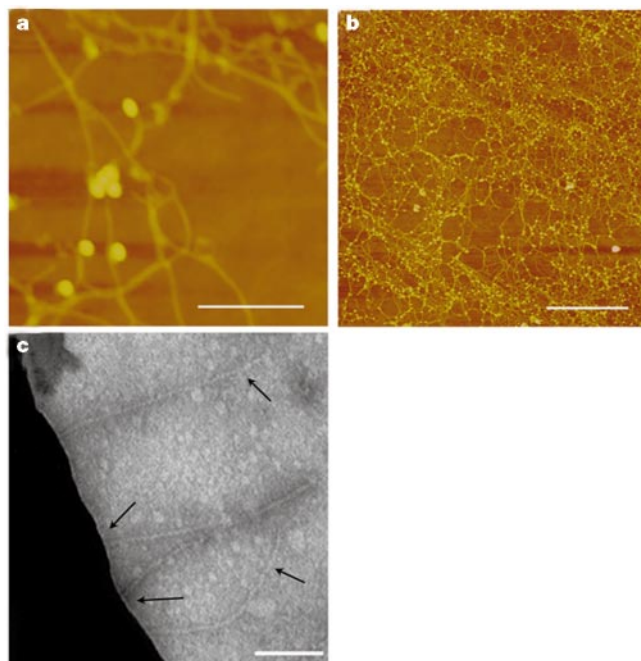


Figure 3 AFM and TEM analysis of peptide–semiconductor recognition. **a, b**, AFM images of G1-3 phage bound to an InP(100) substrate. **a**, Individual phage and their attached Au nanoparticles. Scale bar, 250 nm. **b**, Image showing the uniformity of phage coverage on the InP surface. Scale bar, 2.5 μm. **c**, TEM image of G1-3 phage recognition of GaAs. Individual phage particles are indicated with arrows. Scale bar, 500 nm.

as the thickness of the organic layer increases, the signal from the inorganic substrate decreases. This observation was used to confirm that the intensity of gold was indeed due to the presence of phage on the surface of GaAs.

Binding studies were performed that correlate with the XPS data, where equal numbers of specific phage clones were exposed to various semiconductor substrates with equal surface areas. Wild-type clones (no random peptide insert) did not bind to GaAs (no plaques were detected). For the G1-3 clone, the eluted phage population was 12 times greater from GaAs(100) than from the GaAs(111)A surface.

Using atomic force microscopy (AFM), we imaged the G1-3, G12-3 and G7-4 clones bound to GaAs(100) and InP(100). The InP crystal has a zinc-blende structure, isostructural with GaAs, although the In–P bond has greater ionic character than the Ga–As bond¹³. The 10-nm width and 900-nm length of the observed phage in AFM matches the dimensions of the M13 phage observed by transmission electron microscopy (TEM), and the gold spheres bound to M13 antibodies were observed bound to the phage

(Fig. 3a). As shown in Fig. 3b, the InP surface has a high concentration of phage. These images suggest that there are many factors involved in substrate recognition, including atom size, charge, polarity and crystal structure.

The G1-3 clone (negatively stained) is shown bound to a GaAs crystalline wafer (dark in colour) in the TEM image in Fig. 3c. The arrows indicate individual phage particles, bound through the pIII minor coat protein located at one end of the phage clone. This image confirms that binding was directed by the modified pIII protein of G1-3, not through non-specific interactions with the major coat protein. This illustrates the feasibility of using specific peptide–semiconductor interactions in assembling nanostructures and heterostructures (Fig. 4e).

We used fluorescence microscopy to demonstrate the preferential attachment of phage to a zinc-blende surface in close proximity to a surface of differing chemical and structural composition. A nested square pattern was etched into a GaAs wafer; this pattern contained 1- μm lines of GaAs, and 4- μm SiO₂ spacings in between each line, (Fig. 4a, b). The G12-3 clones were interacted with the GaAs/SiO₂ patterned substrate, washed to reduce non-specific binding, and tagged with an immuno-fluorescent probe, tetramethyl rhodamine (TMR). The tagged phage are seen as the three red lines and the centre dot, in Fig. 4b, corresponding to G12-3 binding only to GaAs. The SiO₂ regions of the pattern remain unbound by phage and are dark in colour. This result was not observed on a control that was not exposed to phage, but which was exposed to the primary antibody and TMR (Fig. 4a).

We observed that the GaAs clone G12-3 was substrate-specific for GaAs over AlGaAs (Fig. 4c). AlAs and GaAs have essentially identical lattice constants at room temperature, 5.66 Å and 5.65 Å, respectively, and thus ternary alloys of Al_xGa_{1-x}As can be epitaxially grown on GaAs substrates. GaAs and AlGaAs have zinc-blende crystal structures, but the G12-3 clone exhibited selectivity in binding only to GaAs. A multilayer substrate was used, consisting of alternating layers of GaAs and of Al_{0.98}Ga_{0.02}As. The substrate material was cleaved and subsequently reacted with the G12-3 clone. The G12-3 clones were labelled with 20-nm gold nanoparticles. Examination by scanning electron microscopy (SEM) shows the alternating layers of GaAs and Al_{0.98}Ga_{0.02}As within the heterostructure (Fig. 4c). X-ray elemental analysis of gallium and aluminium was used to map the gold particles exclusively to the GaAs layers of the heterostructure, demonstrating the high degree of binding specificity for chemical composition. In Fig. 4d, we provide a model for the discrimination of phage for semiconductor heterostructures, as seen in the fluorescence and SEM images (Figs 4a–c).

We have shown the power of using phage-display libraries to identify, develop and amplify binding between organic peptide sequences and inorganic semiconductor substrates. Experimental work remains to be performed, including computer simulations to understand better the nature of the peptide–substrate binding demonstrated by these experiments. We have extended this peptide recognition and specificity of inorganic crystals to other substrates, including GaN, ZnS, CdS, Fe₃O₄ and CaCO₃; these results will be reported elsewhere. We are currently designing bivalent synthetic peptides with two-component recognition (Fig. 4e); such peptides have the potential to direct nanoparticles to specific locations on a semiconductor structure. These organic–inorganic pairs should provide powerful building blocks for the fabrication of a new generation of complex, sophisticated electronic structures. □

Methods

Peptide selection

The library was exposed to the semiconductor crystals in Tris-buffered saline (TBS) containing 0.1% TWEEN-20, to reduce phage–phage interactions on the surface. After rocking for 1 h at room temperature, the surfaces were washed with 10 exposures to Tris-buffered saline, pH 7.5, and increasing TWEEN-20 concentrations from 0.1% to 0.5% (v/v). The phage were eluted from the surface by the addition of glycine–HCl (pH 2.2) for

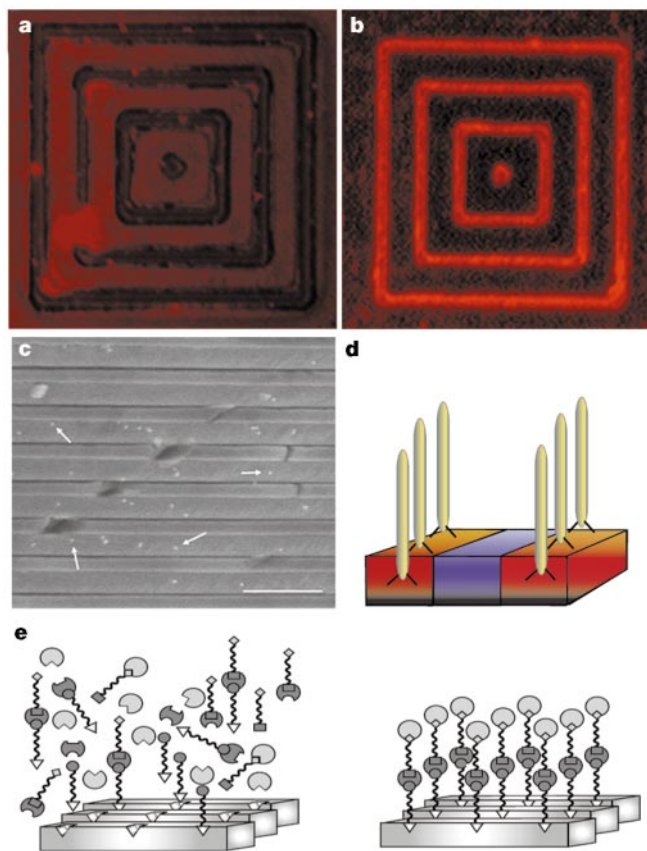


Figure 4 Phage recognition of semiconductor heterostructures. **a–c**, Fluorescence images related to GaAs recognition by phage. **a**, Control experiment: no phage is present, but primary antibody and streptavidin–tetramethyl rhodamine (TMR) are present. **b**, The GaAs clone G12-3 was interacted with a substrate patterned with 1- μm GaAs lines and 4- μm SiO₂ spaces. The phage were then fluorescently labelled with TMR. The G12-3 clone specifically recognized the GaAs and not the SiO₂ surface; scale bar, 4 μm . A diagram of this recognition process is shown in **d**, in which phage specifically attach to one semiconductor rather than another, in a heterostructure. **c**, An SEM image of a heterostructure containing alternating layers of GaAs and Al_{0.98}Ga_{0.02}As, used to demonstrate that this recognition is element-specific. The cleaved surface was interacted with G12-3 phage, and the phage was then tagged with 20-nm gold particles. These nanoparticles (shown arrowed in **c**) are located on GaAs and not AlGaAs layers. Scale bar, 500 nm. **e**, Diagram illustrating the use of this specificity to design nanoparticle heterostructures using proteins with multiple recognition sites.

10 min, transferred to a fresh tube and then neutralized with Tris-HCl (pH 9.1). The eluted phage were titred and binding efficiency was compared.

The phage eluted after third-round substrate exposure were mixed with their *Escherichia coli* ER2537 host and plated on LB XGal/IPTG plates. Since the library phage were derived from the vector M13mp19, which carries the lacZ α gene, phage plaques were blue in colour when plated on media containing Xgal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) and IPTG (isopropyl- β -D-thiogalactoside). Blue/white screening was used to select phage plaques with the random peptide insert. Plaques were picked and DNA sequenced from these plates.

Substrate preparation

Substrate orientations were confirmed by X-ray diffraction, and native oxides were removed by appropriate chemical specific etching. The following etches were tested on GaAs and InP surfaces: NH₄OH: H₂O 1:10, HCl:H₂O 1:10, H₃PO₄: H₂O 3:1:50 at 1 min and 10 min etch times. The best element ratio and least oxide formation (using XPS) for GaAs and InP etched surfaces was achieved using HCl: H₂O for 1 min followed by a deionized water rinse for 1 min. However, since an ammonium hydroxide etch was used for GaAs in the initial screening of the library, this etch was used for all other GaAs substrate experiments. Si(100) wafers were etched in a solution of HF:H₂O 1:40 for one minute, followed by a deionized water rinse. All surfaces were taken directly from the rinse solution and immediately introduced to the phage library. Surfaces of control substrates, not exposed to phage, were characterized and mapped for effectiveness of the etching process and morphology of surfaces by AFM and XPS.

Multilayer substrates of GaAs and of Al_{0.98}Ga_{0.02}As were grown by molecular beam epitaxy onto (100) GaAs. The epitaxially grown layers were Si-doped (n-type) at a level of $5 \times 10^{17} \text{ cm}^{-3}$.

Antibody and gold labelling

For the XPS, SEM and AFM experiments, substrates were exposed to phage for 1 h in Tris-buffered saline then introduced to an anti-fd bacteriophage—biotin conjugate, an antibody to the pIII protein of fd phage, (1:500 in phosphate buffer, Sigma) for 30 min and then rinsed in phosphate buffer. A streptavidin/20-nm colloidal gold label (1:200 in phosphate buffered saline (PBS), Sigma) was attached to the biotin conjugated phage through a biotin—streptavidin interaction; the surfaces were exposed to the label for 30 min and then rinsed several times with PBS.

XPS

The following controls were done for the XPS experiments to ensure that the gold signal seen in XPS was from gold bound to the phage and not non-specific antibody interaction with the GaAs surface. The prepared (100) GaAs surface was exposed to (1) antibody and the streptavidin—gold label, but without phage, (2) G1-3 phage and streptavidin—gold label, but without the antibody, and (3) streptavidin—gold label, without either G1-3 phage or antibody.

The XPS instrument used was a Physical Electronics Phi ESCA 5700 with an aluminium anode producing monochromatic 1,487-eV X-rays. All samples were introduced to the chamber immediately after gold-tagging the phage (as described above) to limit oxidation of the GaAs surfaces, and then pumped overnight at high vacuum to reduce sample outgassing in the XPS chamber.

AFM

The AFM used was a Digital Instruments Bioscope mounted on a Zeiss Axiovert 100s-2tv, operating in tip scanning mode with a G scanner. The images were taken in air using tapping mode. The AFM probes were etched silicon with 125- μm cantilevers and spring constants of 20–100 N m⁻¹ driven near their resonant frequency of 200–400 kHz. Scan rates were of the order of 1–5 $\mu\text{m s}^{-1}$. Images were levelled using a first-order plane fit to remove sample tilt.

TEM

TEM images were taken using a Philips EM208 at 60 kV. The G1-3 phage (diluted 1:100 in TBS) were incubated with GaAs pieces (500 μm) for 30 min, centrifuged to separate particles from unbound phage, rinsed with TBS, and resuspended in TBS. Samples were stained with 2% uranyl acetate.

SEM

The G12-3 phage (diluted 1:100 in TBS) were incubated with a freshly cleaved hetero-structure surface for 30 min and rinsed with TBS. The G12-3 phage were tagged with 20-nm colloidal gold. SEM and elemental mapping images were collected using the Norian detection system mounted on a Hitachi 4700 field emission scanning electron microscope at 5 kV.

Received 16 November 1999; accepted 10 April 2000.

1. Belcher, A. M. *et al.* Control of crystal phase switching and orientation by soluble mollusc-shell proteins. *Nature* **381**, 56–58 (1996).
2. Falini, G. *et al.* Control of aragonite or calcite polymorphism by mollusk shell macromolecules. *Science* **271**, 67–69 (1996).
3. Cha, J. N. Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones *in vitro*. *Proc. Natl Acad. Sci. USA* **96**, 361–365 (1999).
4. Meldrum, F. C., Mann, S., Heywood, B. R., Frankel, R. B. & Bazylinski, D. A. Electron microscopy study of magnetosomes in two cultured vibrioid magnetotactic bacteria. *Proc. R. Soc. Lond. B* **251**, 238–242 (1993).

5. Colvin, V. L., Goldstein, A. N. & Alivisatos, A. P. Semiconductor nanocrystals covalently bound to metal surfaces with self-assembled monolayers. *J. Am. Chem. Soc.* **114**, 5221–5230 (1992).
6. Brust, M., Bethell, D., Schiffrin, D. J. & Kiely, C. J. Novel gold-dithiol nano-networks with nonmetal electronic properties. *Adv. Mater.* **7**, 795–797 (1995).
7. Li, M., Wong, K. K. W. & Mann, S. Organization of inorganic nanoparticles using biotin-streptavidin connectors. *Chem. Mater.* **11**, 23–26 (1999).
8. Alivisatos, A. P. *et al.* Organization of 'nanocrystal molecules' using DNA. *Nature* **382**, 609–611 (1996).
9. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
10. Brown, S. Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage λ receptor. *Proc. Natl Acad. Sci. USA* **89**, 8651–8655 (1992).
11. Brown, S. Metal-recognition by repeating polypeptides. *Nature Biotechnol.* **15**, 269–272 (1997).
12. Parmley, S. F. & Smith, G. P. Antibody-selectable filamentous Fd phage vectors—affinity purification of target genes. *Gene* **73**, 305–318 (1988).
13. Swaminathan, V. & Macrander, A. T. *Materials Aspects of GaAs and InP Based Structures* (Prentice Hall, Englewood Cliffs, New Jersey, 1991).

Acknowledgements

We thank D. Margolese, D. Morse and G. Stucky for discussions, and H. Reese, J. English and R. Naone for providing semiconductor substrates. We acknowledge the use of the core microscopy facilities in the Texas Materials Institute (SEM). We also thank the Institute of Molecular and Cellular Biology (TEM) at the University of Texas at Austin. We acknowledge the assistance of the NSF-sponsored National Nanofabrication Users Network in providing some of the structures for this project. This work was supported by ARO/DARPA (S.R.W. and A.M.B.), a DuPont Young Investigator Award (A.M.B.), the NSF (P.F.B. and E.L.H.) and the Robert A. Welch Foundation (P.F.B. and A.M.B.). This work was also funded by faculty start-up funds provided by the University of Texas at Austin (A.M.B.).

Correspondence and requests for materials should be addressed to A.M.B. (e-mail: belcher@mail.utexas.edu).

Reduced growth of Alaskan white spruce in the twentieth century from temperature-induced drought stress

Valerie A. Barber*†‡, Glenn Patrick Juday†‡ & Bruce P. Finney*

* Institute of Marine Science, † Forest Sciences Department, University of Alaska Fairbanks, Fairbanks, AK 99775, USA

‡ These authors contributed equally to the work

The extension of growing season at high northern latitudes seems increasingly clear from satellite observations of vegetation extent and duration^{1,2}. This extension is also thought to explain the observed increase in amplitude of seasonal variations in atmospheric CO₂ concentration. Increased plant respiration and photosynthesis both correlate well with increases in temperature this century and are therefore the most probable link between the vegetation and CO₂ observations³. From these observations^{1,2}, it has been suggested that increases in temperature have stimulated carbon uptake in high latitudes^{1,2} and for the boreal forest system as a whole⁴. Here we present multi-proxy tree-ring data (ring width, maximum late-wood density and carbon-isotope composition) from 20 productive stands of white spruce in the interior of Alaska. The tree-ring records show a strong and consistent relationship over the past 90 years and indicate that, in contrast with earlier predictions, radial growth has decreased with increasing temperature. Our data show that temperature-induced drought stress has disproportionately affected the most rapidly growing white spruce, suggesting that, under recent climate warming, drought may have been an important factor limiting carbon uptake in a large portion of the North American boreal forest. If this limitation in growth due to drought stress is sustained, the future capacity of northern latitudes to sequester carbon may be less than currently expected.

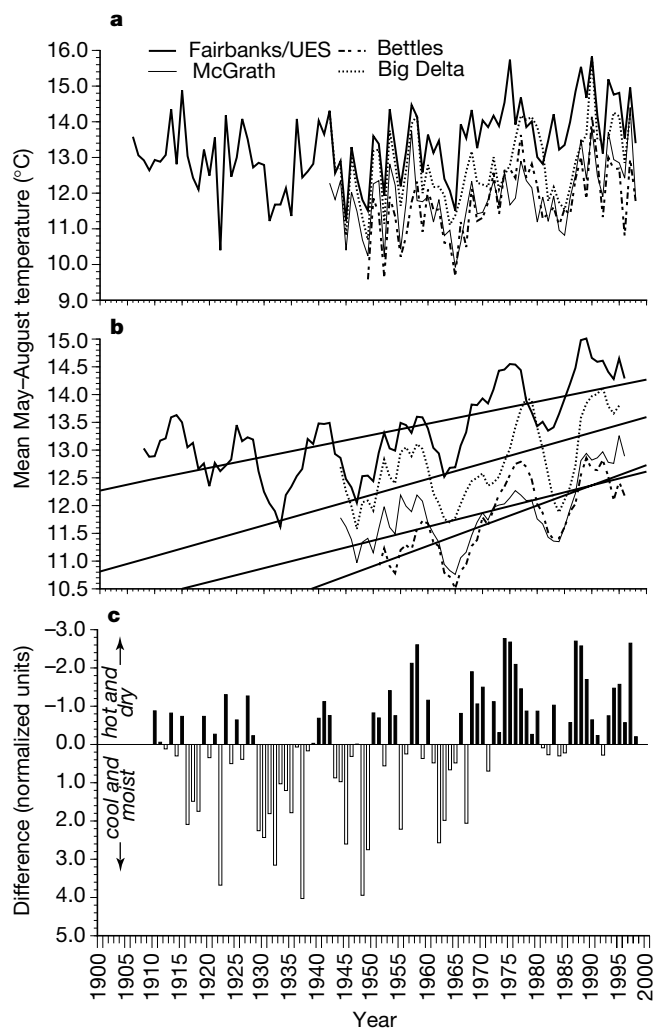


Figure 1 Climate trends in interior Alaska in the twentieth century. Mean summer (May through August) temperature at four meteorological stations (see Fig. 2 for locations).

a, annual values; **b**, annual values smoothed with five-year running mean and fitted with

regression line; **c**, normalized growth year (September through August) precipitation minus normalized May through August temperature at Fairbanks. Scaled to zero mean over the period 1906–98.

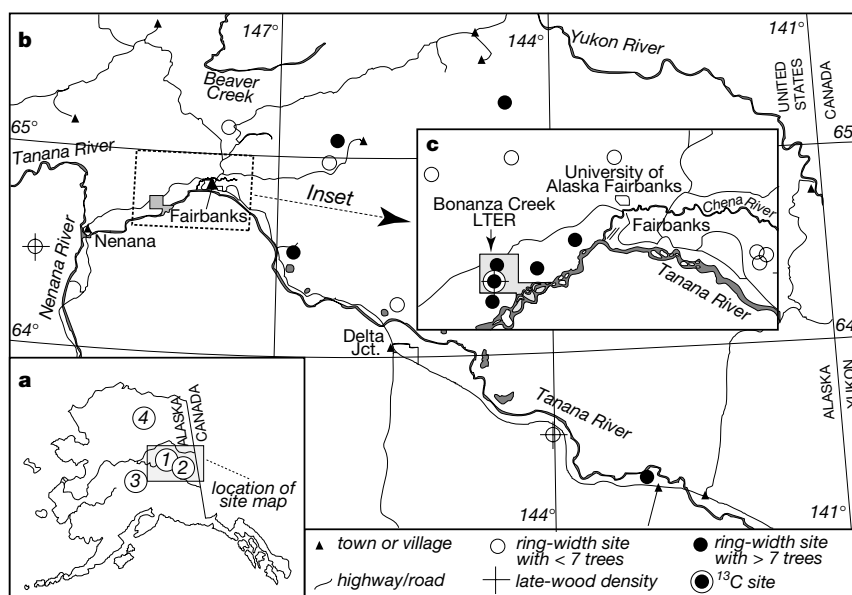


Figure 2 Map of field area. **a**, Location of meteorological stations. Numbered circles represent meteorological stations: 1, Fairbanks/University Experiment Station; 2, Big Delta; 3, McGrath; 4, Bettles. **b**, Extensive tree-ring sampling sites in east-central Alaska.

c, Inset of intensive tree-ring sampling locations in and near Bonanza Creek Long-term Ecological Research (LTER). See legend for **b** and **c**.

We assembled the available twentieth-century summer climate records from stations representative of the boreal forest region of Alaska (Fig. 1a, b). Although the greatest magnitude of high latitude warming has been reported in winter and early spring⁵, these Alaska stations show a strong warming trend in the growing season over the past 50 years (Fig. 1a, b). Interior Alaska is semi-arid with potential evapotranspiration equal to annual precipitation at many interior locations⁶. A combined index of growing season temperature and growth-year precipitation in central Alaska registers severe and prolonged warmth and dryness from the 1970s to the present that is unprecedented in the twentieth century (Fig. 1c). Warming without a concurrent increase in precipitation is projected to turn regions of the present-day central Canadian boreal forest into lower productivity Aspen parklands⁷.

We sampled trees from a broad range of diameters from closed-canopy white spruce stands in the Bonanza Creek Long-Term Ecological Research site and surrounding areas across east-central Alaska (Fig. 2). The sample represents mature and old stands (Fig. 3a) representative of the Alaskan boreal forest, in contrast to the well studied forest-tundra treeline^{8–10}.

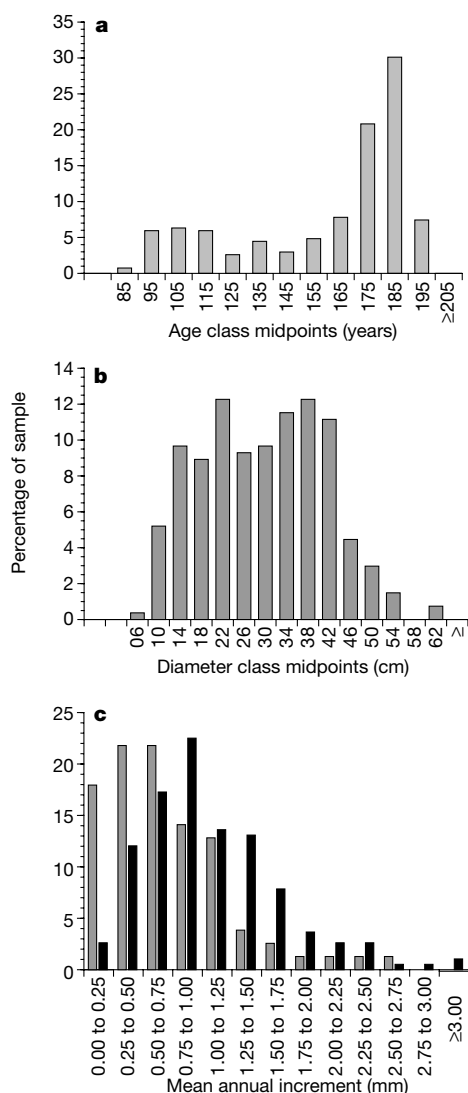


Figure 3 Characteristics of white spruce radial-growth sample. **a**, **b**, Minimum tree age (**a**) and diameter (**b**), $n = 269$ trees. **c**, Frequency distribution of twentieth-century radial growth for trees correlated ($p < 0.05$) with Fairbanks ring-width climate index ($n = 191$) (black bars) and trees not correlated ($p > 0.05$) with ring-width climate index ($n = 78$) (grey bars).

Correlation analyses of all three tree-ring parameters for 36 months (year of tree-ring formation plus the two years before) with the Fairbanks/University Experiment Station record show highly significant correlation with summer temperatures (Fig. 4a–c). We did not use other interior Alaska climate stations because they are highly correlated with the Fairbanks record (Fig. 1) and are shorter in duration. Ring-width is strongly negatively correlated with summer monthly mean temperature in the year of ring formation plus the two years before (Fig. 4a). The negative relationship of ring-width and summer temperature contrasts with the positive relationship widely reported for forest-tundra margin trees^{8,10,11}. Growth-year precipitation (September through August) is correlated positively ($p > 0.01$) with ring-width both for the year of ring formation (0.34 annual values, 0.64 for five-year running mean (smoothed)) and the year before (0.39 annual values, 0.68 smoothed). Most conifers are determinate growers: photosynthetic gain in the growth season before ring formation has the strongest influence on current-year growth¹². By contrast, the $\delta^{13}\text{C}$ and maximum late-wood density (MLWD) properties in spruce ring-width are correlated most strongly with mean monthly summer temperatures of the current growth year (Fig. 4b, c).

We combined monthly mean summer temperature and growth year precipitation into an index of climatic favourability for white spruce growth (see Methods). High radial growth and favourable climate occur consistently during the 50-year period 1915–65 (Fig. 5a). The ring-width climate index (CI_{rw}) closely mirrors one- and two-year positive and negative growth anomalies in the sample (Fig. 5a), such as the spike of growth in response to a cool and moist 1937 and the short, sharp reduction of growth in

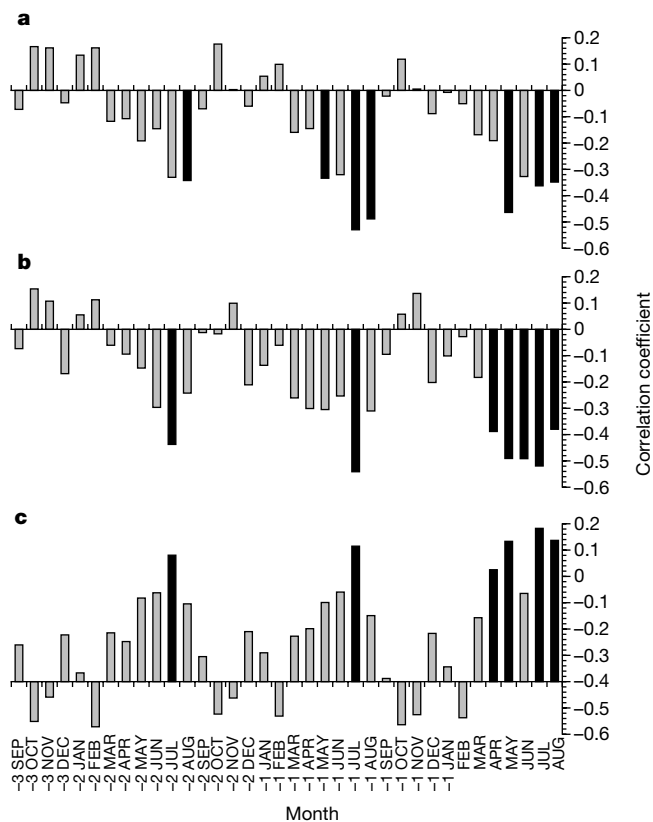


Figure 4 Correlation of tree-ring properties to Fairbanks mean monthly temperature for the three years before completion of tree-ring growth. The year is indicated by a number from -1 to -3 . All values are Pearson correlation coefficients; correlations significant at $p < 0.01$ are shaded black. **a**, Ring width; **b**, $\delta^{13}\text{C}$ discrimination; **c**, maximum late-wood density. Correlation values are for the period 1909–1996 (**a** and **b**) and 1918–94 (**c**).

response to the warm and dry years of 1957–58 (Fig. 1c). A quasi-decadal cyclicity is apparent in which opposite trends in mean radial growth occur regularly over periods of 5–6 years (Fig. 5a, b). The most striking feature of the climate and ring-width record is the sustained unfavourable climate signal and growth response from the 1970s to the present.

Several features of the sample indicate that this climatically related growth reduction is significant at the landscape level of carbon uptake. Mean ring-width of each of the 20 stands and radial growth of a majority (71%) of the individual trees in the sample are significantly correlated ($p > 0.05$) with CI_{rw} . The population of trees that is significantly correlated with climate displays a distribution skewed towards faster rates of growth in the twentieth century than the portion of the population that is not significantly correlated (Fig. 3c). Thus the fastest growing trees are most limited by summer warmth and low precipitation.

Correlation analyses of the $\delta^{13}C$ and wood density properties provide independent confirmation of the ring-width results. The carbon-isotope ($\delta^{13}C$) ratio provides information on CO_2 uptake and water vapour loss during photosynthesis¹³, and thus under limiting conditions can register drought stress. We calculated discrimination of $\delta^{13}C$, the difference between $\delta^{13}C$ of the tree-ring (hollocellulose) and atmospheric $\delta^{13}C$ annually. The $\delta^{13}C$ discrimination of our sampled white spruce is significantly negatively correlated with summer monthly mean temperatures in the

year of the ring formation (Fig. 4b).

Fairbanks precipitation is significantly correlated with $\delta^{13}C$ discrimination (0.335, $p > 0.02$), but combining summer temperature with growth-year precipitation did not result in a significantly higher overall correlation than the correlation with summer temperature alone. The $\delta^{13}C$ discrimination of white spruce displays a strong relationship to the short-term variability of Fairbanks summer temperature throughout the twentieth century (Fig. 5c) and declines in the 1980s and 1990s to an unprecedented and sustained low level for the twentieth century (Fig. 5d).

MLWD is the maximum density of the wood formed at the end of the growing season when conifer tracheids switch from large-diameter, thin-walled cells (formed in spring and early summer) to smaller-diameter, thick-walled cells. Correlation analysis has shown that this number integrates growing conditions over the length of the photosynthetically active season¹⁴ and is also more consistent than ring-width¹⁵. We hypothesize that an increase in the length of the growing season along with warmer summer temperatures depletes already limiting soil moisture. In an environment of high soil-moisture tension, wood cell production shifts earlier in the summer from relatively low-density early wood to higher-density late wood, prolonging the wall-thickness phase of growth which is associated with increased late-wood density^{16,17}. The MLWD of our white spruce sample is significantly correlated with early and late summer temperatures in the year of ring formation (Fig. 4c), as

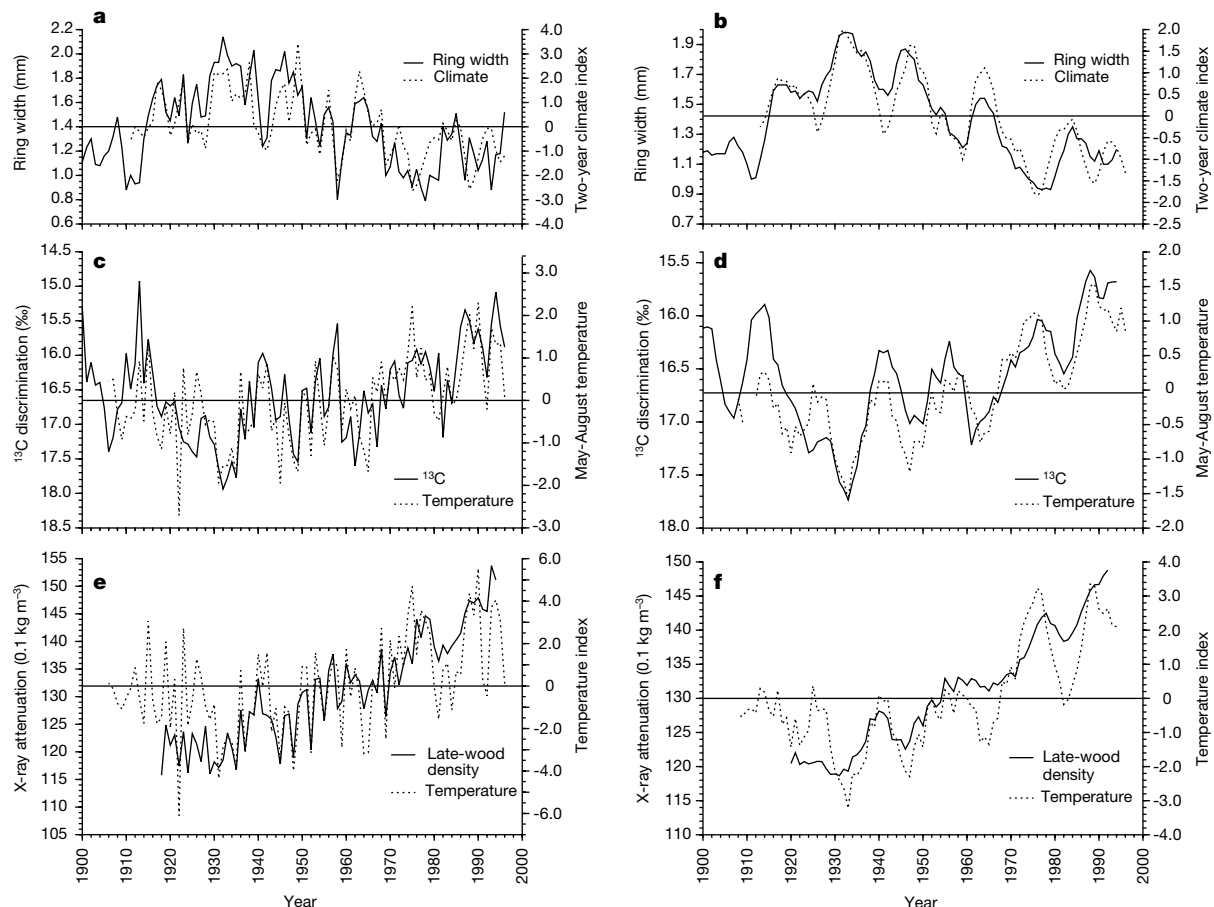


Figure 5 Tree-ring properties in relation to Fairbanks climate during the twentieth century. Annual values on left, smoothed (five-year running mean) on right. All climate values are normalized with zero mean and scaled as standard deviation units. All correlations are significant at $p < 0.001$. **a**, Ring-width versus two-year average Fairbanks climate index (see Methods for definition of ring-width climate index; see Fig. 1c for one-year values of ring-width climate index), correlation = 0.75. **b**, Smoothed ring-

width and climate index, correlation = 0.91. **c**, $\delta^{13}C$ discrimination versus May through August temperature, correlation = 0.69. **d**, Smoothed $\delta^{13}C$ discrimination and May through August temperature, correlation = 0.85. **e**, Maximum late-wood density versus density climate index (mean of normalized May, July and August temperature), correlation = 0.75. **f**, smoothed maximum late-wood density and density climate index, correlation = 0.92.

noted for North American treeline white spruce¹⁵, and other months add little improvement. An unprecedented late-twentieth-century rise in MLWD and the associated temperature index is consistent with the other tree-ring properties (Fig. 5e, f).

Our carbon isotope and MLWD results confirm that drought stress accounts for the climate sensitivity of ring-width for productive upland white spruce sites in interior Alaska. Climate sensitivity in our sample has been sustained (neither degraded nor improved) over the last nine decades, and drought stress has reached levels unprecedented in the twentieth century because of climate warming experienced in the 1980s and 1990s. The wide distribution (Fig. 2), representative diameter distribution of the sampled tree population (Fig. 3a, b) and consistent response of the stands suggest that this response is typical of upland white spruce forests in interior Alaska. In the western Canadian boreal forest, white spruce radial growth displays a negative relationship with the duration of fire weather (hot and dry)¹⁸, suggesting that our findings have broader applicability. Our results are also consistent with the suggestion that drought stress is one possible explanation for a late-twentieth-century decrease in the positive relationship of temperature and radial growth of trees at the forest–tundra margin across the Northern Hemisphere^{19,20}.

White spruce is one of the most productive and widespread forest types in the boreal forest of western North America^{21–23}. Thus any coherent, climate-related change in white spruce growth is likely to be an important factor in CO₂ uptake in the boreal forest, a region that is one of the planet's major potential carbon sinks. Our results suggest that the current assumption of carbon-cycle models²⁴, of a uniform positive relationship of atmospheric carbon uptake to high-latitude warming, will lead to an overestimate of high-latitude carbon storage and an underestimate of future atmospheric CO₂. □

Methods

The radial-growth sample is representative of trees in mature and old stands that are dominant on the contemporary landscape, including trees across a broad range of diameters. The radial-growth sample includes 269 trees in 20 separate stands; radial growth was calculated as a stand-weighted mean. All trees contributed from 1908 until the date of sampling (sampling took place between 1994 and 1996) or, in two cases, the death of the stand. One stand (Reserve West in Bonanza Creek) was killed by wildfire in 1983 and the other was logged in 1987. In these two stands wood disks were harvested from the stumps. Density was measured in cores from 23 trees collected in three stands including Reserve West and outlying sites to the east and west (Fig. 2). The common period of analysis for density was 1918–1994; all trees contributed throughout, except that sample depth in one stand dropped from 14 to 6 trees between 1982 and 1994. Stable carbon isotope ($\delta^{13}\text{C}$) was measured only at Reserve West. Four orthogonal cores from each of four trees pooled together give an isotope value representative of a site²⁵. We obtained isotope measurements for the period 1900–1981 from four orthogonal wedges from each of four harvested stump disks. For the period 1982–96 samples were obtained from four cores collected at each of four surviving trees near the fire perimeter. We believe the isotope sample is representative of white spruce of the interior Alaskan boreal forest because Reserve West is centrally located, is within the well studied Long-term Ecological Research site, and the tree-ring parameters are well correlated with each other. We confirmed the trends in $\delta^{13}\text{C}$ seen at Reserve West in limited samples covering three and four decades collected at two other sites located about 30 km from Bonanza Creek.

We did not transform ring widths into de-trended normalized ring-width index values for two reasons. First, we wanted to preserve the weighting effect of tree increments of different sizes because this reflects annual production more closely than normalized values, which remove this effect. Second, most white spruce of the age range in this study exhibit little age-related growth trend, and the removal of trend is user-specified and risks the loss of information on long-term climate variability that may actually influence tree growth²⁶. However, we also correlated climate with de-trended ring-width indices²⁷ for several stands and found little difference, demonstrating that our results are not an artefact of age-related changes in ring width.

We examined various combinations of climate parameters in order to produce climate indices that maximized correlation for each of the three tree-ring properties (Fig. 5). The climate index (CI_{rw}) that correlates best with ring-width (Fig. 5a, b) is composed of the following parameters:

$$\text{CI}_{\text{rw}} = (\text{Yr}_0(\text{PPT}_{\text{gr}})_n - (T_{\text{M}-\text{A}})_n) + \text{Yr}_{-1}(\text{PPT}_{\text{gr}})_n - (T_{\text{M}-\text{A}})_n)/2$$

where Yr₀ is the year of ring formation, Yr₋₁ is the year prior to ring formation, (PPT_{gr})_n is the normalized growth year precipitation, and (T_{M-A})_n is the normalized May–August temperature.

For stable carbon isotope analysis, four orthogonal samples of annual rings were

excised, ground and blended with the same-year wood from four tree disks. Holocellulose was extracted from the annual wood²⁸ and carbon-isotope ratios were analysed for the years 1900–81. The same procedure was used on cores from the surviving trees for the period 1967–96, to provide an overlap period for correction. Isotope trends for the period of 1967–81 follow similar curves from both sets of wood samples, but there is a slight offset (0.7) in the $\delta^{13}\text{C}$ of the fire-killed trees versus the live trees, probably due to site-specific effects. A correction was made for this offset. A correction was also made for $\delta^{13}\text{C}$ changes in atmospheric CO₂ over the last 150 years due to fossil fuel and biomass combustion using the Law Dome Antarctic ice core data (1.53‰ over 150 years) and discrimination was calculated²⁹. Thus a continuous record of $\delta^{13}\text{C}$ discrimination was determined for 1900–96 and the best correlation was found with May–August temperatures (Fig. 4b, 5c, d).

Maximum late-wood density was measured by X-ray attenuation^{15,30} at Lamont–Doherty Earth Observatory. One to four radial samples were collected from each tree. Annual density values were first combined to produce annual tree averages that were then used to calculate overall yearly stand means for each of three stands. The three-stand means were then averaged to produce the sample mean used in correlation analysis. We determined that MLWD correlated best with May, July and August temperatures (Fig. 4c, 5e, f).

Received 25 August 1999; accepted 10 April 2000.

- Keeling, C. D., Chin, J. F. S. & Whorf, T. P. Increased activity of northern vegetation inferred from atmospheric CO₂ measurements. *Nature* **382**, 146–149 (1996).
- Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981–1991. *Nature* **386**, 698–702 (1997).
- Chapin, F. S. I., Zimov, S. A., Shaver, G. R. & Hobbie, S. E. CO₂ fluctuation at high latitudes. *Nature* **383**, 585–586 (1996).
- Braswell, B. H., Schimel, D. S., Linder, E. & Moore III, B. The response of global terrestrial ecosystems to interannual temperature variability. *Science* **278**, 870–872 (1997).
- Chapman, W. L. & Walsh, J. E. Recent variations of sea ice and air temperature in high latitudes. *Bull. Am. Meteorol. Soc.* **74**, 33–47 (1993).
- Patric, J. H. & Black, P. E. *Potential Evapotranspiration and Climate in Alaska by Thornthwaite's Classification*. (USDA Forest Service Research Paper PNW-71, Juneau, Alaska, 1968).
- Hogg, E. H. & Hurdle, P. A. The aspen parkland in western Canada: a dry-climate analogue for the future boreal forest? *Wat. Air Soil Pollut.* **82**, 391–400 (1995).
- Jacoby, G. C. & D'Arrigo, R. Reconstructed northern hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America. *Clim. Change* **14**, 39–59 (1989).
- Jacoby, G. C., D'Arrigo, R. D. & Juday, G. P. Tree-ring indicators of climatic change at northern latitudes. *World Res. Rev.* **11**, 21–29 (1999).
- Garfinkle, H. L. & Brubaker, L. B. Modern climate–tree-growth relationships and climatic reconstruction in sub-Arctic Alaska. *Nature* **286**, 872–874 (1980).
- Jacoby, G. C., D'Arrigo, R. D. & Davaajants, T. Mongolian tree rings and 20th-Century warming. *Science* **273**, 771–773 (1996).
- Kozlowski, T. T. & Pallardy, S. G. *Physiology of Woody Plants* 1–411 (Academic Press, San Diego, California, 1997).
- Livingston, N. J. & Spittlehouse, D. L. Carbon isotope fractionation in tree rings early and late wood in relation to intra-growing season water balance. *Plant Cell Environ.* **19**, 768–774 (1996).
- Jacoby, G. C., Ivanciu, I. S. & Ulan, L. D. A 263-year record of summer temperature for northern Quebec reconstructed from tree-ring data and evidence of a major climatic shift in the early 1800's. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **64**, 69–78 (1988).
- D'Arrigo, R. D., Jacoby, G. C. & Free, R. M. Tree-ring width and maximum latewood density at the North American tree line: parameters of climate change. *Can. J. Forest Res.* **22**, 1290–1296 (1992).
- Denne, M. P. Effects of environmental change on wood production and wood structure in *Picea sitchensis* seedlings. *Ann. Bot.* **40**, 1017–1028 (1976).
- Conkey, L. E. Red spruce tree-ring widths and densities in eastern North America as indicators of past climate. *Quat. Res.* **26**, 232–243 (1986).
- Larsen, C. P. S. & MacDonald, G. M. Relations between tree-ring widths, climate, and annual area burned in the boreal forest of Alberta. *Can. J. Forest Res.* **25**, 1746–1755 (1995).
- Jacoby, G. C. & D'Arrigo, R. D. Tree ring width and density evidence of climatic and potential forest change in Alaska. *Glob. Biogeochem. Cycles* **9**, 227–234 (1995).
- Briffa, K. R. *et al.* Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* **391**, 678–682 (1998).
- Elliott-Fisk, D. L. in *North American Terrestrial Vegetation* (eds Barbour, M. C. & Billings, W. D.) 33–62 (Cambridge University Press, Cambridge, 1988).
- Ruess, R. W., Van Cleve, K., Yarie, J. & Viereck, L. A. Comparative estimates of fine root production in successional taiga forests on the Alaskan interior. *Can. J. Forest Res.* **26**, 1326–1336 (1996).
- Van Cleve, K., Chapin III, F. S., Flanagan, P. W., Viereck, L. A. & Dyreness, C. T. *Forest Ecosystems in the Alaskan Taiga: A Synthesis of Structure and Function* 230 (Springer, New York, 1986).
- Kicklighter, D. W. *et al.* A first-order analysis of the potential role of CO₂ fertilization. *Tellus B* **51**, 343–366 (1999).
- Leavitt, S. W. & Long, A. Sampling strategy for stable carbon isotope analysis of tree-rings in pine. *Nature* **311**, 145–147 (1984).
- Briffa, K. R. & Osborn, T. J. Seeing the wood from the trees. *Science* **284**, 926–927 (1999).
- Cook, E. R. & Kairiukstis, L. A. *Methods of Dendrochronology: Applications in the Environmental Sciences* (Kluwer Academic, Dordrecht, The Netherlands, 1990).
- Leavitt, S. W. & Danzer, S. R. Method for batch processing small wood samples to holocellulose for stable-carbon isotope analysis. *Anal. Chem.* **65**, 87–89 (1992).
- Francy, R. J. *et al.* A 1000-year high precision record of $\delta^{13}\text{C}$ in atmospheric CO₂. *Tellus B* **51**, 170–193 (1999).
- Thetford, R. D., D'Arrigo, R. D. & Jacoby, G. C. An image analysis system for determining densitometric and ring-width time series. *Can. J. Forest Res.* **21**, 1544–1548 (1991).

Acknowledgements

This work was carried out with the financial support of the McIntire–Stennis Cooperative

Forestry Research Program, the US NSF Long-Term Ecological Research (LTER) Program and Paleoclimates from Arctic Lakes and Estuaries (PALE) Program, the University of Alaska's Natural Resource Fund, International Arctic Research Centre (IARC), and the Centre for Global Change and Arctic System Research. We thank T. Chapin for enthusiastic support and helpful review of this manuscript. We also thank G. Jacoby for use of his laboratory and equipment for density analyses, and A. Krumhardt for field and laboratory assistance.

Correspondence and requests for materials should be addressed to V.A.B. (e-mail: barber@ims.uaf.edu).

Isotopic evidence for Late Cretaceous plume–ridge interaction at the Hawaiian hotspot

Randall A. Keller*, Martin R. Fisk* & William M. White†

* College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA

† Department of Geological Sciences, Cornell University, Ithaca, New York 14853, USA

When a mantle plume interacts with a mid-ocean ridge, both are noticeably affected. The mid-ocean ridge can display anomalously shallow bathymetry, excess volcanism, thickened crust, asymmetric sea-floor spreading and a plume component in the composition of the ridge basalts^{1–4}. The hotspot-related volcanism can be drawn closer to the ridge, and its geochemical composition can also be affected^{5–7}. Here we present Sr–Nd–Pb isotopic analyses of samples from the next-to-oldest seamount in the Hawaiian hotspot track, the Detroit seamount at 51° N, which show that, 81 Myr ago, the Hawaiian hotspot produced volcanism with an isotopic signature indistinguishable from mid-ocean ridge basalt. This composition is unprecedented in the known volcanism from the Hawaiian hotspot, but is consistent with the interpretation from plate reconstructions⁸ that the hotspot was located close to a mid-ocean ridge about 80 Myr ago. As the rising mantle plume encountered the hot, low-viscosity asthenosphere and hot, thin lithosphere near the spreading centre, it appears to have entrained enough of the isotopically depleted upper mantle to overwhelm the chemical characteristics of the plume itself. The Hawaiian hotspot thus joins the growing list of hotspots that have interacted with a rift early in their history.

For more than 81 Myr (ref. 9), the Hawaiian hotspot has been producing the age-progressive Hawaiian–Emperor chain of volcanic islands and seamounts (Fig. 1). This chain extends for 5,800 km from the present location of volcanic activity on the island of Hawaii and Loihi seamount, to the northernmost Emperor seamount (Meiji) at 53° N. There are over 100 volcanic edifices along the Hawaiian–Emperor chain, and more than 30 of these (in addition to the Hawaiian Islands) have been sampled by dredge and drill^{9–11}, making it the most intensively sampled and studied hotspot track on Earth.

Published Sr–isotope ratios of tholeiitic basalts from along the Hawaiian hotspot track show an interesting trend with age (Fig. 2). These ratios remain fairly constant along the Hawaiian Islands and the Hawaiian ridge between Kilauea volcano on Hawaii and Yuryaku seamount (Fig. 1); they then decrease steadily northwards along the Emperor seamounts to Suiko guyot, which was the oldest seamount previously analysed (64.7 Myr; ref. 12). This decrease in the ⁸⁷Sr/⁸⁶Sr ratio back in time was attributed to a decrease in distance between the hotspot and the nearest spreading centre¹³. Only the tholeiitic basalts (as opposed to the transitional and alkalic

basalts) show this isotopic trend, perhaps because the tholeiites experienced little if any interaction with the lithosphere^{14–16}.

Because this trend in ⁸⁷Sr/⁸⁶Sr with time was all within the range for the Hawaiian Islands, it could be explained by a sampling bias in the older seamounts. We therefore decided to measure the isotopic ratios of basalts from the two oldest Emperor seamounts (Detroit and Meiji) to determine if the isotopic trend continued further back in time. We also analysed a single sample from Suiko to tie in with the published data. We chose to concentrate our analyses, and our discussion here, on the tholeiitic basalts because only they show the isotopic trend with time.

We measured the Sr, Nd and Pb isotope ratios of basalts from two locations (ODP Sites 883 and 884) on the Detroit seamount platform⁹. The Site 884 basalts have low concentrations of alkalis, TiO₂, and other incompatible elements at moderate MgO (4.6–8.8 wt%) and Ni (47–94 p.p.m.) contents, and are unambiguously tholeiitic⁹. Basalt recovered at the other drillsite on Detroit (Site 883) are too altered to determine unambiguously if they are tholeiitic (and we wish to limit our study to tholeiites); however, we include them here (Table 1) because they are isotopically similar to the Site 884 basalts (Fig. 3), which bolsters our argument that the Late Cretaceous volcanism was significantly different from the more recent volcanism. The Suiko sample from DSDP Site 433 and the Meiji sample from DSDP Site 192 are tholeiitic basalts^{12,17}.

Age-corrected (initial) ⁸⁷Sr/⁸⁶Sr, ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb values (Table 1) in the Detroit tholeiites (Site 884) are lower than anything previously reported from the Hawaiian–Emperor chain (Fig. 3), while ¹⁴³Nd/¹⁴⁴Nd and ²⁰⁶Pb/²⁰⁴Pb values for these tholeiites overlap the Hawaiian field. Age corrections to these isotopic data are small (Table 1), and even the measured ratios are distinct from the Hawaiian data. The Sr and Pb isotope ratios of the Detroit tholeiites are within the range of Pacific mid-ocean-ridge basalt (MORB), while their ¹⁴³Nd/¹⁴⁴Nd values are lower than modern Pacific MORB, but within the range of Pacific MORB data adjusted to 80 Myr (Fig. 3b). Isotopic data for the Site 883 samples are similar to the Site 884 data, but are not quite as depleted.

The Suiko sample has the highest Sr and Pb isotope ratios measured in this study, and values of ¹⁴³Nd/¹⁴⁴Nd that are similar to those found at Meiji and Detroit (Table 1). Suiko isotopic ratios

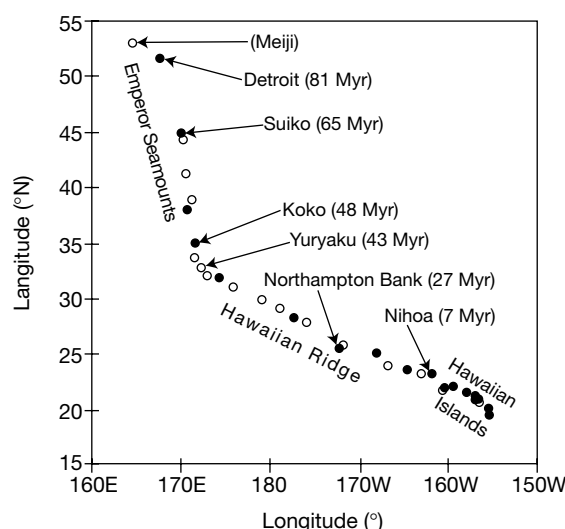


Figure 1 Location of volcanoes in the Hawaiian–Emperor island/seamount chain that have been dated (the age of Meiji is unknown). Additional locations have been sampled, but not dated. Locations from which tholeiites have been reported are shown by filled circles. Locations where no tholeiites were reported, or the identification of the tholeiites is in doubt, are shown by open circles. Seamounts mentioned in the text are labelled with names and ages. Age data are from a compilation¹⁰, except for the Detroit seamount age⁹. New isotopic data presented in this study are from Suiko, Detroit and Meiji.

are within the range of published values for tholeiites from the Hawaiian Islands and the Hawaiian ridge (although its $^{87}\text{Sr}/^{86}\text{Sr}$ is slightly lower than the lowest published $^{87}\text{Sr}/^{86}\text{Sr}$ for Hawaiian tholeiites that also have Nd and Pb data available; Fig. 3). Our $^{87}\text{Sr}/^{86}\text{Sr}$ value for the Suiko sample is within analytical error of previous Suiko tholeiite analyses^{13,18}. Isotopic ratios of the Meiji sample are generally intermediate between the Detroit and Suiko data (Table 1).

The uniqueness of the Cretaceous Emperor seamount isotopic data compared to the rest of the Hawaiian–Emperor chain is most obvious on the plot of $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ (Fig. 3a). The Detroit data overlap the Pacific MORB field, and are distinct from

the Hawaiian Islands/ridge field. The Meiji and Suiko data are displaced from the highly depleted values of the Detroit field, towards the high- $^{206}\text{Pb}/^{204}\text{Pb}$ – low- $^{87}\text{Sr}/^{86}\text{Sr}$ end of the Hawaiian Islands/ridge field. The positive slope of our Meiji–Detroit–Suiko data array on this plot is obviously distinct from the negative slope of the array formed by tholeiites from the 0 to 48 Myr part of the Hawaiian–Emperor chain.

We can quickly rule out several explanations for the unradiogenic nature of the early products of the Hawaiian hotspot. (1) The fact that the Sr and Pb isotope compositions of the Detroit tholeiites are unlike any other known product of the Hawaiian hotspot shows that the trend in Fig. 2, now extended back to 81 Myr, is not due to

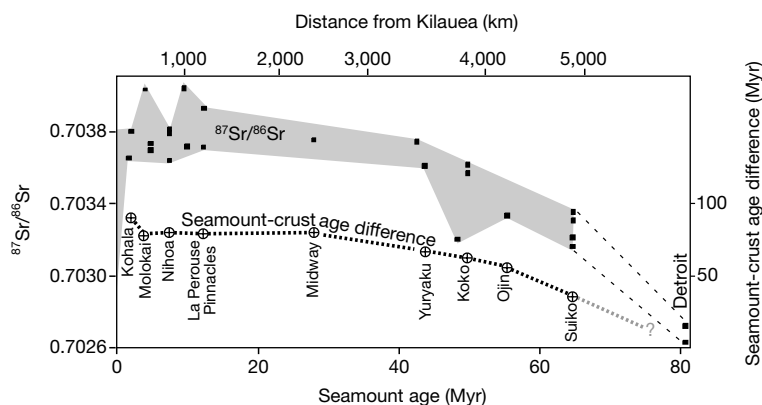


Figure 2 Evolution of the Hawaiian hotspot with time. The shaded field shows the range of published $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of tholeiitic basalts versus seamount age (lower axis) and distance from Kilauea volcano along the Hawaiian–Emperor chain (upper axis). Data points used to define the shaded field are shown by small filled squares, except that the many data points from the Hawaiian Islands have been omitted for clarity. The extension of

the shaded field to include new data presented here for the Detroit seamount is shown by the two dashed lines. The crossed circles connected by the thick dotted line shows the trend in age difference (right axis) between the seamounts and their underlying crust. Age data sources are as in Fig. 1, except that the age difference between the Detroit seamount and the crust is estimated from isochron charts²⁷ to be <20 Myr.

Table 1 Isotopic and trace-element ratios of Emperor seamount basalts

	Seamount					
	Suiko	Detroit	Detroit	Detroit	Detroit	Meiji
Location						
Sample	433-31-1	883-1-3	883-2-3	884-6-3	884-10-4	192-6-2
Age (Myr)	64.7	80?	80?	81.2	81.2	85?
Latitude (N)	44°46.6'	51°11.9'	51°11.9'	51°27.0'	51°27.0'	53°0.0'
Longitude (E)	170°1.2'	167°46.1'	167°46.1'	168°20.2'	168°20.2'	164°42.8'
Depth (m)	2237	3208	3219	4720	4751	4069
Measured ratios						
$^{87}\text{Sr}/^{86}\text{Sr}$	0.703297	0.702884	0.702864	0.702768	0.702652	0.703079
$^{143}\text{Nd}/^{144}\text{Nd}$	0.513068	0.513081	0.513100	0.513184	0.513152	0.513072
$^{206}\text{Pb}/^{204}\text{Pb}$	18.609	18.355	18.500	18.175	18.058	18.750
$^{207}\text{Pb}/^{204}\text{Pb}$	15.463	15.443	15.452	15.419	15.410	15.453
$^{208}\text{Pb}/^{204}\text{Pb}$	38.136	37.856	37.809	37.680	37.582	38.036
Rb/Sr	0.015	0.024	0.024	0.012	0.010	0.031
Sm/Nd	0.278	0.285	0.285	0.349	0.350	0.280
Th/Pb	0.833	0.941	0.938	0.321	0.312	0.812
U/Pb	0.300	0.470	0.438	0.143	0.156	0.500
La/Sm	1.99	1.78	1.86	1.06	1.08	1.99
Initial ratios						
$^{87}\text{Sr}/^{86}\text{Sr}$	0.703257	0.702808	0.702788	0.702730	0.702620	0.702972
$^{143}\text{Nd}/^{144}\text{Nd}$	0.512994	0.512987	0.513006	0.513068	0.513035	0.512974
$^{206}\text{Pb}/^{204}\text{Pb}$	18.394	17.938	18.113	18.047	17.918	18.279
$^{207}\text{Pb}/^{204}\text{Pb}$	15.453	15.423	15.434	15.413	15.403	15.431
$^{208}\text{Pb}/^{204}\text{Pb}$	37.945	37.589	37.543	37.588	37.492	37.791

All isotopic ratios were measured at Cornell University using published procedures²⁵. All samples were repeatedly leached in hot HCl before digestion, and all isotopic ratios were corrected for fractionation. Sr and Pb isotope ratios are normalized to standards NBS987 (= 0.710248) and NBS981 (= 16.937; 15.493; 36.705), respectively. The average $^{143}\text{Nd}/^{144}\text{Nd}$ for the La Jolla Nd standard measured at the time of these analyses (via Ames Nd) was 0.511876 ± 14 . Two-sigma standard errors based on numerous analyses of standards during the time of the unknown runs are: $^{87}\text{Sr}/^{86}\text{Sr} \pm 0.000014$, $^{143}\text{Nd}/^{144}\text{Nd} \pm 0.000016$, $^{206}\text{Pb}/^{204}\text{Pb} \pm 0.010$, $^{207}\text{Pb}/^{204}\text{Pb} \pm 0.011$, and $^{208}\text{Pb}/^{204}\text{Pb} \pm 0.030$. Within-run standard errors are smaller than these external errors. Trace-element concentrations were measured by isotope dilution thermal ionization mass spectrometry (TIMS-ID) of leached samples (Pb, Th and U) and inductively coupled plasma mass spectrometry (ICP-MS), of unleached samples (Rb, Sr, La, Nd and Sm). The ages of the Site 433 and Site 884 samples are known from ^{40}Ar – ^{39}Ar dating^{10,13}. For age correction purposes, the Site 883 samples are assumed to be 1 Myr younger than the Site 884 samples because Site 883 was slightly higher on the seamount, and the Site 883 samples appear to be of transitional composition. The Site 192 sample is estimated to be 85 Myr old for age correction purposes. Errors in the estimated ages of the Site 192 and Site 883 samples of several million years or less are not relevant here.

sampling bias. The Hawaiian Islands are well sampled, and it is unlikely that anything as depleted as the Detroit tholeiites has escaped notice. (2) The low isotopic ratios are not an artefact of the age-correction procedure: age-corrected Sr and Pb isotope ratios for the Site 884 tholeiites are barely different from the measured values. Also, the two Detroit locations (Sites 883 and 884) have similar age-corrected isotopic ratios, despite their different measured ratios and trace-element compositions. (3) The shift in isotopic characteristics of the Hawaiian hotspot from 81 Myr to the present cannot be explained by "plume ageing", that is, radioactive decay within the plume and its source region, as has been proposed for the Kerguelen plume¹⁹. The trace-element ratios required to account for the differences between the Detroit isotopic data and the Hawaiian Islands/ridge data ($Rb/Sr = 0.278$, $U/Pb = 0.558-2.11$, $Th/Pb = 2.09$) are 4 to 28 times these ratios in the Site 884 tholeiites (and due to partial melting effects, these ratios should be even lower in the source than in the basalts), and 1 to 2 orders of magnitude higher than the maximum ratios likely to be found in plumes²⁰.

We interpret the displacement of the Cretaceous data points towards MORB values as mixing between a depleted component with the isotopic characteristics of MORB, and a 'Kilauea-like' plume component at the low- $^{87}Sr/^{86}Sr$ end of the Hawaiian Islands/ridge data set. There is no evidence in the isotopic data from any of the Late Cretaceous samples for the high- $^{87}Sr/^{86}Sr$, low- $^{143}Nd/^{144}Nd$ and low- $^{206}Pb/^{204}Pb$ ('Koolau-like') component found in the modern plume.

Plate reconstructions place an active spreading centre close to the position of the Hawaiian hotspot at about 80 Myr ago (ref. 8). The rising plume must have entrained more depleted mantle when it came up near the spreading centre. In other locations where a spreading ridge is close to a hotspot (for example, Bouvet, Easter, Galápagos and Iceland) the isotopic characteristics of the hotspot

volcanism extend toward MORB-like values^{2,21,22}. A nearby spreading ridge could have created a higher-temperature and lower viscosity and density regime in the upper mantle at ~81 Myr ago, conditions that lead to more entrainment²². Hotter upper-mantle temperatures close to the spreading centre also could have allowed greater degrees of melting of the upper mantle²³, and thus more entrainment. The nearby spreading centre could have caused more vigorous upper-mantle convection and a mechanical stirring effect. Also, younger, hotter lithosphere may be more readily assimilated by the ascending plume melts, thereby imparting a MORB-like isotopic signature to the hotspot basalts. The thick dotted line connecting crossed circles in Fig. 2 shows the changes in the age relationship between the seamounts and their underlying crust. Older lithosphere (at present ~90 Myr at the Hawaiian Islands) is colder, less easily melted, and is thought to play an insignificant role in the genesis of tholeiites on the Hawaiian Islands¹⁴⁻¹⁶. The thickness (and thus age) of the lithosphere could also determine how much the asthenosphere contributes to hotspot volcanism. If the upwelling mantle plume begins to melt when it reaches a depth of, say, approximately 100 km, and continues to produce melt until it reaches the base of the lithosphere, an area of very young (thin) lithosphere will have a taller melting zone that melts more asthenosphere.

Some workers have suggested that the head of a new mantle plume could entrain large amounts of depleted upper mantle²³⁻²⁵ (although numerical models suggest that the melting zone of a plume head entrains less rather than more upper mantle²⁶), but it is not known if this model can be applied to the Hawaiian plume. This is because the old end of the Hawaiian hotspot track ends at a subduction zone, so the existence of older seamounts or a plume-head-associated flood basalt is speculative.

Finally, a change in the isotopic characteristics of the source region of the plume cannot be ruled out. The region of the mantle

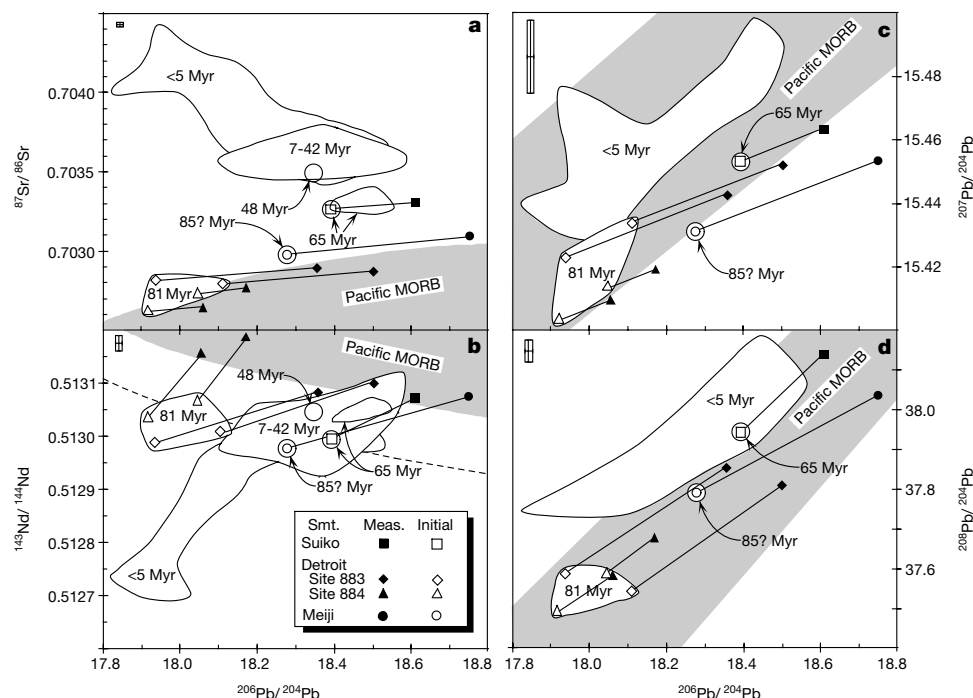


Figure 3 Isotope ratios for tholeiitic basalts. The figure shows plots of the $^{206}Pb/^{204}Pb$ ratio against other isotopic ratios; data are for tholeiitic basalts from the Hawaiian Islands (<5 Myr field), the Hawaiian ridge (7–42 Myr field), Koko guyot (48 Myr), and Suiko (65 Myr), Detroit (81 Myr) and Meiji (~85? Myr) seamounts. For Suiko, Detroit and Meiji, filled symbols are measured ratios and open symbols are age-corrected (initial) ratios (Table 1). Only data from tholeiites are shown, with the possible exception of the

Meiji and the Site 883 Detroit samples. Hawaiian Islands and Pacific MORB data are from numerous sources. Fields without data points (7–42 Myr, 48 Myr and 65 Myr) in panels **a** and **b** are also age-corrected¹⁸. The dashed line in panel **b** is the lower bound of the Pacific MORB field adjusted to 80 Myr ago using a MORB-reservoir Sm/Nd ratio of 0.36 (ref. 28). Age adjustments to the MORB fields in the other panels are negligible. Rectangle in the upper left corner of each panel shows $\pm 2\sigma$ error box.

where the Hawaiian plume acquires its isotopic signature has probably not been homogeneous and static. Isotopically diverse packages of mantle material may have been randomly swept into this region over (at least) the past 81 Myr. The more depleted composition of the Detroit samples could be explained by a package of relatively depleted mantle having been swept into the plume source region during the Cretaceous. However, this model is untestable, and does not explain why the other two hotspots from which long-term data are available (Kerguelen and Réunion) also appear to have become more enriched with time^{19,23–25}. □

Received 3 December 1999; accepted 17 April 2000.

1. Schilling, J.-G. Iceland mantle plume: Geochemical evidence along Reykjanes Ridge. *Nature* **242**, 565–571 (1973).
2. Verma, S. P., Schilling, J.-G. & Wagoner, D. G. Neodymium isotopic evidence for Galapagos hotspot-spreading centre system evolution. *Nature* **306**, 654–657 (1983).
3. Ito, G. & Lin, J. Oceanic spreading center-hotspot interactions: Constraints from along-isochron bathymetric and gravity anomalies. *Geology* **23**, 657–660 (1995).
4. Muller, R. D., Roest, W. R. & Royer, J.-Y. Asymmetric sea-floor spreading caused by ridge-plume interactions. *Nature* **396**, 455–459 (1998).
5. White, W. M., McBirney, A. R. & Duncan, R. A. Petrology and geochemistry of the Galapagos Islands: Portrait of a pathological mantle plume. *J. Geophys. Res.* **98**, 19533–19563 (1993).
6. Kincaid, C., Ito, G. & Gable, C. Laboratory investigation of the interaction of off-axis mantle plumes and spreading centres. *Nature* **376**, 758–761 (1995).
7. Kingsley, R. H. & Schilling, J.-G. Plume-ridge interaction in the Easter-Salas y Gomez seamount chain-Easter microplate system: Pb isotope evidence. *J. Geophys. Res.* **103**, 24159–24177 (1998).
8. Mammerickx, J. & Sharman, G. F. Tectonic evolution of the North Pacific during the Cretaceous quiet period. *J. Geophys. Res.* **93**, 3009–3024 (1988).
9. Keller, R. A., Duncan, R. A. & Fisk, M. R. Geochemistry and ⁴⁰Ar/³⁹Ar geochronology of basalts from ODP Leg 145 (North Pacific Transect). *Proc. ODP Sci. Res.* **145**, 333–344 (1995).
10. Clague, D. A. & Dalrymple, G. B. The Hawaiian-Emperor volcanic chain, Part I, Geologic evolution. *US Geol. Surv. Prof. Pap.* **1350**, 5–54 (1987).
11. Lonsdale, P., Dieu, J. & Natland, J. Posterosional volcanism in the Cretaceous part of the Hawaiian hotspot trail. *J. Geophys. Res.* **98**, 4081–4098 (1993).
12. Dalrymple, G. B., Lanphere, M. A. & Clague, D. A. Conventional and ⁴⁰Ar/³⁹Ar K-Ar ages of volcanic rocks from Ojin (Site 430), Nintoku (Site 432), and Suiko (Site 433) seamounts and the chronology of volcanic propagation along the Hawaii-Emperor chain. *Init. Rep. DSDP* **55**, 659–676 (1980).
13. Lanphere, M. A., Dalrymple, G. B. & Clague, D. A. Rb-Sr systematics of basalts from the Hawaii-Emperor volcanic chain. *Init. Rep. DSDP* **55**, 695–706 (1980).
14. Chen, C.-Y. & Frey, F. A. Trace element and isotopic geochemistry of lavas from Haleakala volcano, East Maui, Hawaii: Implications for the origin of Hawaiian basalts. *J. Geophys. Res.* **90**, 8743–8768 (1985).
15. Hauri, E. H. Major element variability in the Hawaiian mantle plume. *Nature* **382**, 415–419 (1996).
16. Garcia, M. O., Ito, E., Eiler, J. M. & Pietruszka, A. J. Crustal contamination of Kilauea volcano magmas revealed by oxygen isotope analyses of glass and olivine from Puu Oo eruption lavas. *J. Petrol.* **39**, 803–817 (1998).
17. Dalrymple, G. B., Lanphere, M. A. & Natland, J. H. K-Ar minimum age for Meiji Guyot, Emperor seamount chain. *Init. Rep. DSDP* **55**, 677–683 (1980).
18. Unruh, D. M., Stille, P. & Tatsumoto, M. Pb, Sr, and Nd isotopic study of basalts from the Hawaiian-Emperor chain (abstr.) in *Hawaiian Symposium on How Volcanoes Work* (eds Decker, R., Halbig, J., Hazlett, R., Okamura, R. & Wright, T.) 260 (Hawaii Volcano Observatory, Hilo, 1987).
19. Class, C., Goldstein, S. L., Galer, S. J. G. & Weis, D. Young formation age of a mantle plume source. *Nature* **362**, 715–721 (1993).
20. Weaver, B. L. The origin of ocean island basalt end-member compositions: trace element and isotopic constraints. *Earth Planet. Sci. Lett.* **104**, 381–397 (1991).
21. Hanan, B. B. & Schilling, J.-G. Easter microplate evolution; Pb isotope evidence. *J. Geophys. Res.* **94**, 7432–7448 (1989).
22. Hauri, E. H., Whitehead, J. A. & Hart, S. R. Fluid dynamic and geochemical aspects of entrainment in mantle plumes. *J. Geophys. Res.* **99**, 24275–24300 (1994).
23. Weis, D. et al. in *The Indian Ocean: A Synthesis of Results from the Ocean Drilling Program* (eds Duncan, R. A., Rea, D. K., Kidd, R. B., von Rad, U. & Weissel, J. K.) 57–89 (Geophysical Monograph 70, American Geophysical Union, Washington, DC, 1992).
24. Fisk, M. R. et al. Reunion hotspot magma chemistry over the past 65 m.y.: Results from Leg 115 of the Ocean Drilling Program. *Geology* **17**, 934–937 (1989).
25. White, W. M., Cheatham, M. M. & Duncan, R. A. Isotope geochemistry of Leg 115 basalts and inferences on the history of the Reunion mantle plume. *Proc. ODP Sci. Res.* **115**, 53–61 (1990).
26. Farnetani, D. G. & Richards, M. A. Thermal entrainment and melting in mantle plumes. *Earth Planet. Sci. Lett.* **136**, 251–267 (1995).
27. Renkin, M. L. & Sclater, J. G. Depth and age in the North Pacific. *J. Geophys. Res.* **93**, 2919–2935 (1988).
28. Wood, D. A., Joron, J. L. & Treuil, M. A re-appraisal of the use of trace elements to classify and discriminate between magma series erupted in different tectonic settings. *Earth Planet. Sci. Lett.* **45**, 326–336 (1979).

Acknowledgements

We thank M. Cheatham for assistance with the TIMS analyses, A. Ungerer for assistance with the ICP-MS analyses, and B. Hanan and E. Hauri for comments on the manuscript. DSDP samples from Meiji and Suiko seamounts were provided by the ODP West Coast Repository. This work was supported by a JOI/USSAC Ocean Drilling Fellowship.

Correspondence should be addressed to R.A.K. (e-mail: rkeller@oce.orst.edu).

Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit

Birger Rasmussen

Department of Geology and Geophysics, University of Western Australia, Nedlands, Western Australia 6907, Australia

The record of Archaean microfossils is sparse¹. Of the few *bona fide* fossil assemblages, most are from shallow-water settings, and they are typically associated with laminated, stromatolitic sedimentary rocks^{2–4}. Microfossils from deep-sea hydrothermal systems have not been reported in Precambrian rocks (> 544 million years old), although thermophilic microbes are ubiquitous in modern sea-floor hydrothermal settings^{5,6}, and apparently have the most ancient lineages^{7,8}. Here, I report the discovery of pyritic filaments, the probable fossil remains of thread-like microorganisms, in a 3,235-million-year-old deep-sea volcanogenic massive sulphide deposit from the Pilbara Craton of Australia. From their mode of occurrence, the micro-organisms were probably thermophilic chemotrophic prokaryotes, which inhabited sub-sea-floor hydrothermal environments. They represent the first fossil evidence for microbial life in a Precambrian submarine thermal spring system, and extend the known range of submarine hydrothermal biota by more than 2,700 million years⁹. Such environments may have hosted the first living systems on Earth, consistent with proposals for a thermophilic origin of life^{10–13}.

The Pilbara Craton of Australia contains one of the most complete sections of well preserved Archaean volcano-sedimentary rocks and is the site of several previous fossil discoveries^{4,14}. From this region, the probable fossil remains of microorganisms were identified in a volcanogenic massive sulphide (VMS) deposit at Sulphur Springs. The deposit is located in low-strain, very low metamorphic grade (prehnite–pumpellyite facies) rocks of the northern Soanesville belt in the eastern Pilbara Craton (Fig. 1a), and displays exceptional textural and structural preservation despite its great age¹⁵. Mineralization is associated with syn-volcanic faults and lies below a regionally extensive unit of silicified sedimentary and volcanoclastic rocks ('marker chert') which caps a succession of mafic, intermediate and felsic volcanic rocks and syn-volcanic dacitic intrusions (Fig. 1b). These rocks are intruded by the ~3,235-Myr-old co-magmatic Strelley granite, which is synchronous with hydrothermal mineralization^{16–18}. The succession was deposited in a deep marine setting, with water depths probably exceeding 1,000 m, as indicated by volcanic textures, sedimentary facies and style of metal precipitation^{15,16}.

The deposit (UTM grid 728976594) consists of stratabound lenses of pyrite, sphalerite, chalcopryrite, tennantite, galena, barite and quartz, and an underlying stringer zone with veins of quartz, pyrite, chalcopryrite, and minor sphalerite and carbonate. The lower part is characterized by massive, granular, honeycomb and filigree sulphide textures, whereas the upper part contains dendritic, colloform and pellet textures¹⁵. Bitumen occurs in small quantities throughout the deposit as blocky fragments and films that are intergrown with sulphides, and oil is present as fluid inclusions in hydrothermal barite¹⁹. The deposit displays an upward and outward zonation from Cu to Zn–Pb, a diagnostic feature of VMS-style mineralization. Galena from the deposit gives a Pb–Pb model age of ~3,260 Myr (ref. 15). The composition of the ore indicates that fluid temperatures probably reached 300 °C (ref. 15), but distal areas were probably much cooler. Much of the mineralization formed by

replacement of volcanic and volcanoclastic rocks close to the sea floor¹⁶, although it has been argued that open-space sulphide textures indicate a setting resembling black-smoker chimneys at hydrothermal vents¹⁵.

Polished thin sections from the deposit were examined by optical microscopy (using transmitted and reflected light) and scanning electron microscopy. Filaments are located in the siliceous cores and bands of colloform structures (Fig. 2a, b) from samples in the upper part of the deposit. These rocks consist of pyrite and microcrystalline quartz (chert), with minor sphalerite, chalcopyrite, tennantite, arsenopyrite, galena, chlorite and carbonate, and up to 10% of silicified volcanic rock fragments. The colloform structures range in size from <0.1 mm to several centimetres, and include botryoidal, coalesced and branching forms. The structures are almost entirely composed of inclusion-rich pyrite, and are enclosed by paragenetically late chert. Much of the pyrite has been recrystallized to massive pyrite by the influx of late-stage hydrothermal fluids, and many of the structures show evidence of progressive replacement by later sulphides.

The banding and concentric textures preserved in many colloform structures are defined by inclusions after the incomplete replacement of original minerals by pyrite (Fig. 2c). In the few samples where original minerals are preserved, they comprise inclusion-rich chert, coarse quartz crystals and rare carbonate rhombs. Although most of the microcrystalline silica has been recrystallized to coarser quartz, some fine morphological detail is preserved as banding. Individual laminae are typically 5–50 µm thick, and are defined by alternating light and dark zones and by trails of irregular pyrite crystals that extend across interlocking quartz crystals. The colloform cores also contain later infilling aggregates of siderite, inclusion-free quartz and chlorite. The colloform structures are cut by fibrous quartz veins that also contain minor chlorite, carbonate, sphalerite and pyrite, which extend up to a few millimetres into the surrounding pyrite (Fig. 2a).

The filaments are located in the paragenetically early chert and coarse-grained quartz, and are absent in later generation silica, carbonate, chlorite and sulphide. The filaments are threadlike,

unbranched and of uniform thickness along their length (0.5–2 µm in diameter, up to 300 µm long). They include straight, sinuous and sharply curved morphologies which are intertwined where they are densely distributed (Fig. 3a–f). They are abundant, with over 300 filaments in a 5 mm × 10 mm area in one polished thin section (~30 µm thick). In reflected light, the filaments appear to be almost entirely composed of many sub-micrometre-sized pyrite crystals. In the cores of colloform structures, most filaments appear to be randomly oriented, but towards the outer margins they are aligned parallel to concentric layering (Fig. 3g) or, less commonly, are oriented sub-perpendicular to the banding (Figs 3h, 4). In some bands (<1 mm to 20 mm long) most filaments lie parallel to lamination (Fig. 4). The filaments extend across the interlocking crystal boundaries of early chert, coarse quartz and carbonate rhombs, and in places are disrupted and displaced by silica-filled penetrative fractures (Fig. 3f).

The filaments can be traced through different focal levels within thin sections, and are present in several thin sections of the same sample and in many samples from various parts of the deposit. Because the samples are from subsurface (300–350 m) diamond drill-core, well below the water table and with no evidence of weathering, emplacement by modern surface fluids can be dismissed. Thus, the filaments are indigenous to the host rocks. The presence of filaments within paragenetically early minerals which are almost completely replaced by later sulphides and their disruption by cross-cutting fractures indicates that they are syngenetic and were lithified before the main stage of sulphide mineralization, coincident with granitoid intrusion 3,235 Myr ago¹⁸.

A biogenic origin is inferred for the filaments from their sinuous morphology, length-wise uniformity and intertwined habit, which are comparable to those of Archaean and younger microfossils^{1–4,20}. They are entirely dissimilar to abiogenic filamentous microstructures formed by the displacement of carbonaceous matter during crystal growth or the movement of ambient pyrite crystals^{21,22}. The limited range of morphotypes may reflect selective diagenetic preservation²³ or low initial diversity. Perhaps most significantly, the changes in preferred orientation of filaments in different areas of

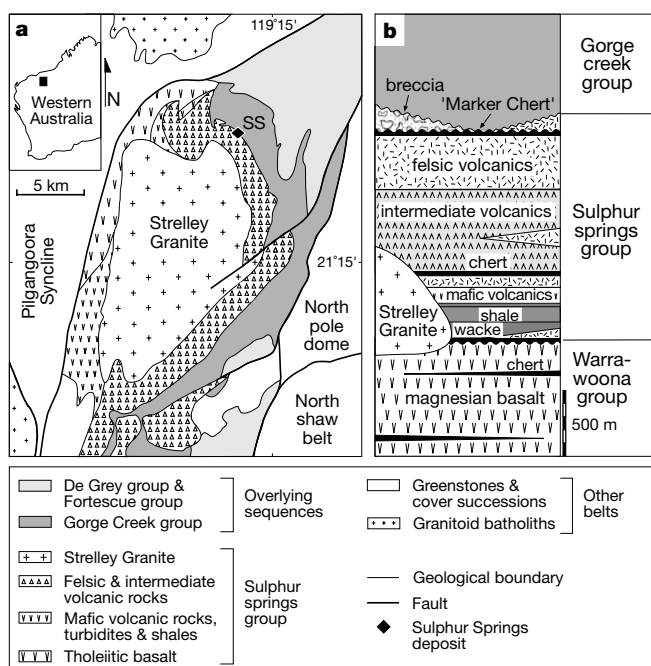


Figure 1 Location and geology of the Sulphur Springs deposit. **a**, Map showing the geology of the northern Soanesville belt and the location of the Sulphur Springs deposit (after ref. 16). **b**, Stratigraphic column of the Sulphur Springs group (after ref. 18).

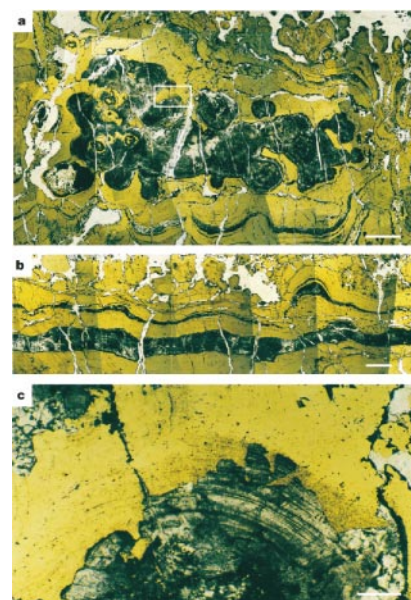


Figure 2 Remnant cores and bands containing filaments. **a**, Silica-rich colloform cores containing filaments, surrounded by replacive pyrite (gold). Combined transmitted and reflected light (TL and RL) photographic montage. **b**, Thin banded layer containing filaments within paragenetically early chert (TL and RL). **c**, Detail of **a** showing the partial replacement of quartz-rich colloform structures displaying fine-scale lamination (TL and RL). Scale bar is 1 mm for **a**, **b** and 0.1 mm for **c**.

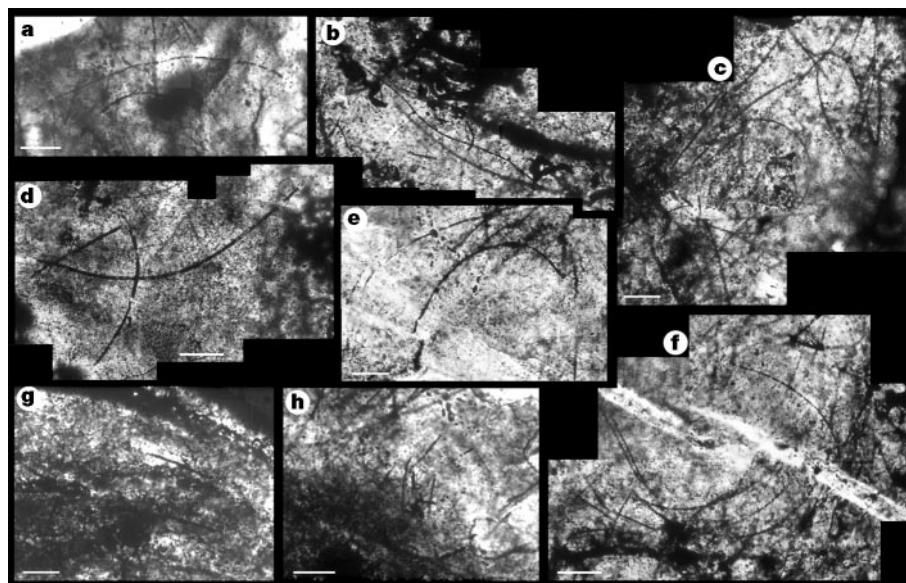


Figure 3 Photomicrographs of filaments from the Sulphur Springs VMS deposit. Scale bar, 10 μm . **a–f**, Straight, sinuous and curved morphologies, some densely intertwined.

g, Filaments parallel to the concentric layering. **h**, Filaments oriented sub-perpendicular to banding.

the colloform structures may indicate behavioural variations in different micro-environments, a distinctly biological feature. Similar patterns are evident in modern microbial mats²⁴ and hot-spring communities²⁵, indicating a biotic tropism or taxis in response to a gradient or fluctuation in a salient ecological parameter. It therefore seems reasonable to infer that the Sulphur Springs filaments are probably biogenic, and as such, represent the best available evidence for microbial life in a hydrothermal setting during the Precambrian.

Although it is difficult to ascribe metabolic characteristics to fossilized microbes on the basis of morphology^{1,2}, some information can be inferred from their mode of occurrence. In the Sulphur Springs deposit, the presence of volcanic rock fragments in many of the fossiliferous samples and their location below a silicified cap of sedimentary and volcanoclastic rocks indicate that the microorganisms probably lived in the pores and crevices of rocks at shallow depths below the sea floor. In such a setting, proximal hydrothermal fluids would have delivered a near-continuous supply of nutrients and metals essential for growth¹². The filaments, which probably acted as nucleation sites^{25,26}, were consequently engulfed by paragenetically early silica, which precipitated as mineral-charged hydrothermal fluids cooled upon approaching the sea floor. As the hydrothermal system evolved, with the formation of hotter fluids owing to the emplacement of the Strelley Granite, the early, low-temperature silica was progressively replaced by pyrite and later sulphides. Although there were several episodes of silicification,

with resultant cementation of pores and replacement of precursor sedimentary and volcanoclastic rocks, it is unclear whether the early silica was generated during volcanism, syn-volcanic intrusion or the earliest stage of granite intrusion.

Given that the filamentous microfossils are encased in hydrothermal silica, and are within a VMS deposit, it seems likely that the microbiota were thermophiles, inhabiting 'low-temperature' (<110 °C) sub-sea-floor niches adjacent to discharging hydrothermal fluids. The oil in the deposit¹⁹ indicates that hydrocarbons entrained by circulating fluids may have provided readily bioavailable energy and carbon sources for microorganisms living around the hydrothermal system^{6,27,28}. The great water depth (>1,000 m) and sediment cover would have offered protection from ultraviolet radiation, and imply that the filamentous microbiota could not have derived energy for growth through light-dependent processes; instead they must have used chemical pathways for metabolic processes. The abundance of reduced chemical species in sea-floor hydrothermal fluids and the probable absence of dissolved oxygen in the Archaean deep ocean²⁹ indicates that these pathways were probably anaerobic. As in modern deep-sea hydrothermal systems^{5,6}, the abundance of reduced sulphur in the hydrothermal fluids and the encrustation of filaments by pyrite indicate that sulphur may have been central to energy-yielding reactions for microbial growth.

Although modern deep-sea hydrothermal systems host diverse microbial communities^{5,6}, reports of microfossils are rare for

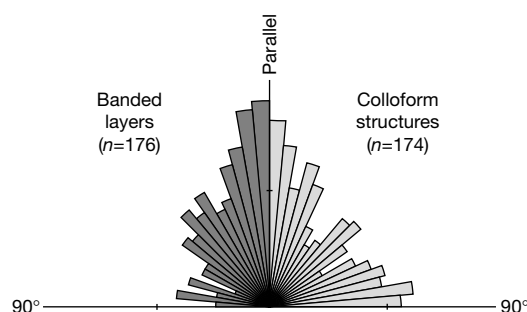


Figure 4 Rose diagram showing filament orientation relative to lamination for colloform structures and banded layers using 5° subdivisions.

ancient deposits^{9,30} and nonexistent for Archaean rocks. Indeed, the oldest fossil microorganisms reported are filamentous bacteria from Cambrian–Ordovician silica–iron oxide exhalites from northeastern Australia⁹. The discovery of probable microfossils in the Sulphur Springs VMS deposit suggests that a chemotrophic deep-sea hydrothermal biosphere thrived over 3,235 Myr ago, some 2,700 Myr before previously described assemblages⁹. Given the likely abundance of submarine thermal springs on the early Earth¹², it is possible that microbial communities flourished in similar settings well before the Sulphur Springs microorganisms, consistent with proposals for a thermophilic origin of life^{7,8,12,13} in deep-sea hydrothermal environments^{10,11}. □

Received 3 December 1999; accepted 6 April 2000.

- Schopf, J. W. in *The Proterozoic Biosphere* (eds Schopf, J. W. & Klein, C.) 25–39 (Cambridge Univ. Press, New York, 1992).
- Schopf, J. W. & Walter, M. R. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 214–239 (Princeton Univ. Press, Princeton, 1983).
- Walsh, M. M. & Lowe, D. R. Filamentous microfossils from the 3,500-Myr-old Onverwacht Group, Barberton Mountain Land, South Africa. *Nature* **314**, 530–532 (1985).
- Awramik, S. M., Schopf, J. W. & Walter, M. R. Filamentous fossil bacteria from the Archean of Western Australia. *Precamb. Res.* **20**, 357–374 (1983).
- Jannasch, H. W. & Mottl, M. J. Geomicrobiology of deep-sea hydrothermal vents. *Science* **229**, 717–725 (1985).
- Karl, D. M. *The Microbiology of Deep-Sea Hydrothermal Vents* (CRC, Boca Raton, 1995).
- Kandler, O. in *Early Life on Earth* Nobel Symposium 84 (ed. Bengtson, S.) 152–160 (Columbia, New York, 1994).
- Stetter, K. O. in *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)* Ciba Foundation Symposium 202 (eds Bock, G. R. & Goode, J. A.) 1–10 (Wiley, Chichester, 1996).
- Duhig, N. C., Davidson, G. J. & Stolz, J. Microbial involvement in the formation of Cambrian sea-floor silica–iron oxide deposits, Australia. *Geology* **20**, 511–514 (1992).
- Corliss, J. B., Baross, J. A. & Hoffman, S. E. An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. *Oceanol. Acta* **4**, 59–69 (1981).
- Russell, M. J. & Hall, A. J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond.* **154**, 377–402 (1997).
- Nisbet, E. G. & Fowler, C. M. R. in *Tectonic, Magmatic, Hydrothermal and Biological Segmentation of Mid-Ocean Ridges* Geological Society Special Publication 118 (eds MacLeod, C. J., Tyler, P. A. & Walker, C. L.) 239–251 (The Geological Society, London, 1996).
- Huber, C. & Wächtershäuser, G. Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: Implications for the origin of life. *Science* **281**, 670–672 (1998).
- Schopf, J. W. & Packer, B. M. Early Archaean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* **237**, 70–73 (1987).
- Vearncombe, S. et al. 3.26 Ga black smoker-type mineralization in the Strelley Belt, Pilbara Craton, Western Australia. *J. Geol. Soc. Lond.* **152**, 587–590 (1995).
- Morant, P. The Panorama Zn–Cu VMS deposits, Western Australia. *Bull. Aust. Inst. Geosci.* **16**, 75–84 (1995).
- Brauhart, C. W., Groves, D. I. & Morant, P. Regional alteration systems associated with volcanogenic massive sulphide mineralisation at Panorama, Pilbara, Western Australia. *Econ. Geol.* **93**, 292–302 (1998).
- Buick, R. et al. Geochronology of the Sulphur Springs Group and Strelley Granite: a temporally distinct igneous province in the Archaean Pilbara Craton, Australia. *Precamb. Res.* (submitted).
- Rasmussen, B. & Buick, R. Oily old ores: Evidence for hydrothermal petroleum generation in an Archaean volcanogenic massive sulphide deposit. *Geology* (in the press).
- Trewin, N. H. & Knoll, A. H. Preservation of Devonian chemotrophic filamentous bacteria in calcite veins. *Palaio* **14**, 288–294 (1999).
- Knoll, A. H. & Barghoorn, E. S. Ambient pyrite in Precambrian chert: New evidence and a theory. *Proc. Natl Acad. Sci. USA* **71**, 2329–2331 (1974).
- Buick, R. Microfossil recognition in Archaean rocks: An appraisal of spheroids and filaments from a 3500 m.y. old chert-barite unit at North Pole, Western Australia. *Palaio* **5**, 441–459 (1991).
- Knoll, A. H., Strother, P. K. & Rossi, S. Distribution and diagenesis of microfossils from the Lower Proterozoic Duck Creek Dolomite, Western Australia. *Precamb. Res.* **38**, 257–279 (1988).
- Bauld, J., D'Amelio, E. & Farmer, J. D. in *The Proterozoic Biosphere* (eds Schopf, J. W. & Klein, C.) 261–269 (Cambridge Univ. Press, New York, 1992).
- Jones, B., Renaut, R. W. & Rosen, M. R. Biogenicity of silica precipitation around geysers and hot spring vents, North Island, New Zealand. *J. Sed. Res.* **67**, 88–104 (1997).
- Ferris, F. G., Beveridge, T. J. & Fyfe, W. S. Iron-silica crystallite nucleation by bacteria in geothermal sediment. *Nature* **320**, 609–611 (1986).
- Bazylynski, D. A., Wirsén, C. O. & Jannasch, H. W. Microbial utilization of naturally-occurring hydrocarbons at the Guaymas Basin hydrothermal vent site. *Appl. Environ. Microbiol.* **55**, 2832–2836 (1989).
- Rueter, P. et al. Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria. *Nature* **372**, 455–458 (1994).
- Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
- Juniper, S. K. & Fouquet, Y. Filamentous iron-silica deposits from modern and ancient hydrothermal sites. *Can. Mineral.* **26**, 859–869 (1988).

Acknowledgements

I thank T. S. Blake, R. Buick and S. Sheppard for comments and discussion; M. G. Doepel and P. Morant for access to samples; and J. Backhouse, B. David, M. G. Doyle, G. L. England, S. Folkert, P. Morant, S. Reverts, S. Richter and K.-H. Wyrwoll for

assistance. This work was supported by an ARC fellowship.

Correspondence and requests for material should be addressed to B.R. (e-mail: brasmuss@geol.uwa.edu.au).

Parasite adaptation to locally common host genotypes

Curtis M. Lively & Mark F. Dybdahl*

Department of Biology, Indiana University, Bloomington, Indiana 47405-3700, USA

* Present address: Department of Biological Sciences, Ohio University, Athens, Ohio, 45701, USA

According to the Red Queen hypothesis—which states that interactions among species (such as hosts and parasites) lead to constant natural selection for adaptation and counter-adaptation—the disproportionate evolutionary success of parasites on common host genotypes leads to correlated selection for sexual reproduction^{1–8} and local adaptation by the parasite population^{9–14}. Here we determined whether local adaptation is due to disproportionate infection of common host genotypes, and, if so, whether infection of common host genotypes is due to commonness *per se*, or some other aspect of these genotypes. In a reciprocal cross-inoculation experiment parasites occupying the same geographical area (sympatric) infected locally common host genotypes significantly more often than rare host genotypes, whereas parasites occupying separate geographical areas (allopatric) showed no such significant difference. A mixed source of parasites (containing F₁ hybrids) also showed no difference in infection between rare and common host genotypes. These results show that local adaptation results from parasite tracking of locally common host genotypes, and, as such, a necessary condition of the Red Queen hypothesis is met.

The Prosobranch snail, *Potamopyrgus antipodarum*, is widely distributed throughout the freshwater habitats of New Zealand, where populations are composed of either obligately asexual females or mixtures of obligately sexual and obligately asexual individuals¹⁵. In a study of a clonal population (Lake Poerua, South Island, New Zealand), we found that common clonal genotypes were more susceptible to infection by a digenetic trematode (*Microphallus* sp.) than were rare clonal genotypes¹⁶. This result is consistent with a theory concerning host–parasite coevolution^{1–8}; nevertheless, alternative explanations are possible. For example, suppose that common clones have a greater competitive ability, but a trade-off between competitive ability and resistance to infection renders them more susceptible to parasites. With this model, the best competitors would be most common, but also most susceptible to infection, and this might explain our previous finding that the most common clones were also the most susceptible to infection by the local source of parasites¹⁶. Here we test the ‘trade-off hypothesis’ by comparing the susceptibility to infection of rare and common clones when exposed to sympatric, allopatric and hybrid sources of parasites. The trade-off hypothesis would predict that the most common clones from Lake Poerua would be significantly more infected than rare clones, independent of the source of parasites. We contrast this prediction with that of the ‘coevolution hypothesis’, which predicts that common host genotypes would be more susceptible to only the sympatric source of parasites.

We exposed individuals from two snail populations (Lake Poerua and Lake Ianthe) to two pure sources of parasites (*Microphallus* sp.)

from the same two lakes. We chose the asexual snail population from Lake Poerua because it is composed of multiple clones. We have determined the relative frequency of these clones using allozyme electrophoresis since 1992. We found that between 1992 and 1996 four clones were common (>15%) in one or more years¹⁶. The remaining fraction of the population (~50% for each year) comprised more than 100 rare clonal genotypes. We chose Lake Ianthe for the allopatric source of parasites. This lake (~80 km south of Lake Poerua) contains a mixed population of sexual and asexual snails¹⁷. The host genotypes that are common in Lake Poerua are not found in Lake Ianthe, and the two snail populations are genetically differentiated on the basis of allozyme frequency data (Nei's genetic distance equal to 0.121; ref. 17). The parasite populations, however, are not genetically differentiated, also on the basis of allozyme data (Nei's genetic distance equal to 0.011; ref. 17). Finally, we exposed snails from both lakes to a 'mixed' source of parasite eggs by combining adult parasites from both lake populations in the same final host (see Methods). Assuming random mating, roughly half the eggs produced should be hybrids of the two parasite sources.

The specific prediction of the trade-off hypothesis was that common clones from Lake Poerua would be over-infected by parasites drawn from both pure sources (Lake Poerua and Lake Ianthe) and the mixed source. In contrast, the coevolution hypothesis predicts (1) that the pure sources of parasites would be better at infecting snails from their sympatric populations (local adaptation) as a result of tracking locally common hosts, and (2) that common clones in Lake Poerua would be over-infected by only the Lake Poerua source of parasites.

The results were consistent with both predictions of the coevolution hypothesis. Parasites were significantly more infective in sympatric host populations than were allopatric sources of parasites and mixed sources of parasites (local adaptation, Fig. 1). In addition, parasites from Lake Poerua infected common sympatric host clones significantly more than rare sympatric host clones (Fig. 2a). In contrast, allopatric (Lake Ianthe) parasites infected rare clones and common clones from Lake Poerua at statistically indistinguishable rates (Fig. 2b), thereby falsifying the trade-off hypothesis. Similarly, the mixed source of parasites also infected rare and common clones at statistically indistinguishable levels, although infection rates of common clones were slightly lower than rare clones (Fig. 2c). This later result suggests that alleles from Lake Ianthe may have disrupted the specific genetic match between Lake Poerua parasites and the host genotypes that were recently common in Lake Poerua.

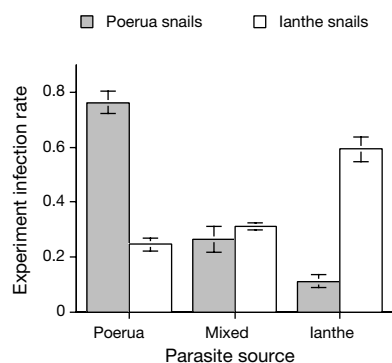


Figure 1 Experimental infection rates of the two snail populations by sympatric, allopatric and mixed parasites. Local parasite adaptation is indicated by significantly higher infection rates for sympatric parasite–host combinations ($\bar{x} = 0.678$; $n = 2$) compared with non-sympatric combinations ($\bar{x} = 0.234$; $n = 4$) (analysis of variance: $F_{1,4} = 28.51$; $P = 0.006$). The mean of four replicates for each treatment combination are given; vertical bars show one standard error of the mean.

In summary, the parasite was found to be adapted to infecting local populations of its snail host, which is consistent with a previous study involving three different lake populations of this same host–parasite combination¹³. In addition, the local adaptation observed here is explained by the greater success of sympatric parasites on locally common host genotypes, which is also consistent with a previous study¹⁶. Finally, the success of parasites on locally common host genotypes was due to commonness in the strict sense, rather than a correlated phenotypic feature of these common genotypes. This combination of results suggests that local adaptation results from genetically based local coevolutionary interactions as proposed in the Red Queen hypothesis. □

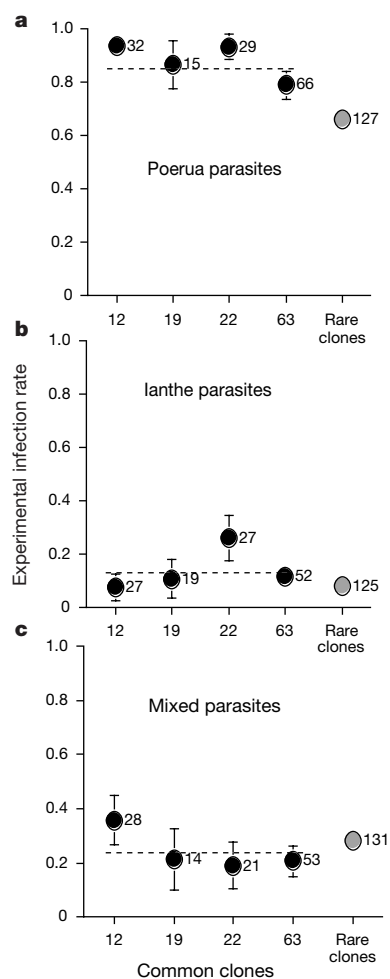


Figure 2 Infection rates of Lake Poerua clones that have been common (12, 19, 22, and 63) and rare between 1992 and 1996 by sympatric, allopatric and mixed parasite sources. Dashed line is the average infection rate of the four common clones. Bars indicate one binomial standard error, and symbol labels give sample size. **a**, Lake Poerua parasites. Common clones were infected at significantly higher rates than rare clones (85.9% and 66.1%, respectively) ($\chi^2 = 14.81$, d.f. = 1, $P < 0.0001$). The four recently common clones were infected at statistically indistinguishable rates ($\chi^2 = 5.935$, d.f. = 3, $P = 0.1148$). **b**, Lake Ianthe parasites. Common clones and rare clones were infected at statistically indistinguishable rates (13.6% and 8.0%, respectively; $\chi^2 = 2.056$, d.f. = 1, $P = 0.1517$). Although the infection rate of clone 22 is slightly higher, the difference in infection rates among the four common clones was not statistically significant ($\chi^2 = 4.267$, d.f. = 3, $P = 0.2341$). When clone 22 is removed, the infection rates of the other three common clones and rare clones are nearly identical (9.5% and 8.0% respectively). **c**, Mixed parasites. Rare clones and common clones were infected at statistically indistinguishable rates (28.2% and 24.1%, respectively; $\chi^2 = 0.537$, d.f. = 1, $P = 0.467$), and the four recently common clones were infected at statistically indistinguishable rates ($\chi^2 = 1.770$, d.f. = 3, $P = 0.6215$).

Methods

Host–parasite system

Potamopyrgus antipodarum serves as the first intermediate host to at least a dozen species of digenetic trematodes. One of these trematode species, *Microphallus* sp., produces encysted larvae (metacercariae) in the snail in 3–4 months under laboratory conditions. The cysts 'hatch' after ingestion by the final host (waterfowl and wading birds), and the resulting hermaphroditic worms produce cross-fertilized eggs within several days; these eggs then pass into the environment. We have found that mice can serve as the final host in laboratory experiments. Snails become exposed to infection after the ingestion of these eggs. An infection resulting from a single egg results in the production of hundreds (or more) asexual larvae within the same snail, thereby sterilizing the host. These larvae then encyst in the snail, but they can be easily removed by dissection.

Experimental infections

Infections of *Potamopyrgus* were carried out in the laboratory (Edward Percival Field Station in Kaikoura, New Zealand) in January 1997 using mice as the final host. Parasite lines were created within 12 laboratory mice by feeding each mouse the metacercarial cysts from 24 infected snails. Four Lake Poerua parasite lines were created using cysts dissected from 24 infected snails collected from the Lake Poerua shoreline; similarly, 4 Lake Ianthe lines were created using cysts from 24 infected snails collected from the Lake Ianthe shoreline. In addition, 4 'mixed' parasite lines were created by combining cysts from 12 infected Lake Poerua snails with 12 infected Lake Ianthe snails. Assuming random mating, 50% of the parasite eggs would be F₁ hybrid genotypes, 25% would be from Poerua × Poerua crosses, and 25% would be from Ianthe × Ianthe crosses. Of the four parasite lines from each parasite source (pure Poerua, pure Ianthe and mixed), two lines were used to infect Lake Poerua snails, and two lines were used to infect Lake Ianthe snails. Parasite eggs were obtained by repeatedly washing the mouse faecal pellets with water. Eggs were collected between two and six days after the mice ingested the cysts.

For each parasite source, we set up 4 replicate containers with 150 snails from the Lake Poerua shoreline and 4 replicate containers with 150 snails from the Lake Ianthe shoreline. Parasite eggs were added to these containers. Snails were kept in the containers with the parasite eggs for 24 days, with water changed twice each day. The snails were then transported to Indiana University where they were held in 4 litres of water. The water was changed regularly and the snails were fed on *Spirulina*. Ninety days after exposure to parasite eggs, we dissected 75 snails from each replicate, and recorded their infection status and the developmental stage of the parasite (early germinal cells lead to blastocercariae, which lead to metacercariae). These stages appear sequentially over a period of about 70–100 days. We limited our analysis to those snails that were infected in the laboratory (that is, early stage infections). We preserved snail tissue samples (head and foot) of Lake Poerua snails for electrophoretic analysis. We used the resulting five-locus allozyme genotypes to identify clonal lineages^{15,16}. We further classified these snails as either 1 of 4 recently common clones (clones 12, 19, 22, 63), which accounted for 50% of the sample, or as rare clones, which contained individuals from 89 different clonal lineages.

Infection rate analysis

For results shown in Fig. 1, statistical analysis was conducted using SPSS¹⁸ on untransformed data, where the dependent variable was mean prevalence of infection for the four replicates within a treatment combination (sympatric versus non-sympatric). The homogeneity-of-variance assumption of the analysis was not violated (Bartlett's Box test: $F_{1,23} = 0.145$; $P = 0.706$).

For results shown in Fig. 2, we used a hierarchical log-linear analysis, with replicate, clone identity and infected as factors. For each of the parasite sources, we examined clone-by-infected interaction terms for common clones and rare clones as groups. Significant interactions would indicate that *Microphallus* sources differentially infected rare versus common clones. We also examined clone-by-infected terms for the four common clones individually to see whether they were differentially infected. We report likelihood ratio χ^2 statistics from a backward model selection routine in SPSS¹⁸. Main and interaction effects involving replicates were non-significant except for the replicate-by-infected term for the mixed source.

Received 6 December 1999; accepted 10 April 2000.

- Jaenike, J. An hypothesis to account for the maintenance of sex within populations. *Evol. Theor.* **3**, 191–194 (1978).
- Bremermann, H. J. Sex and polymorphism as strategies in host–pathogen interactions. *J. Theor. Biol.* **87**, 671–702 (1980).
- Hamilton, W. D. Sex versus non-sex versus parasite. *Oikos* **35**, 282–290 (1980).
- Seger, J. & Hamilton, W. D. in *The Evolution of Sex* (eds Michod, R. E. & Levin, B. R.) 176–193 (Sinauer and Associates, Sunderland, 1988).
- Nee, S. Antagonistic coevolution and the evolution of genotype randomization. *J. Theor. Biol.* **140**, 499–518 (1989).
- Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites (A review). *Proc. Natl Acad. Sci. USA* **87**, 3566–3573 (1990).
- Howard, R. S. & Lively, C. M. Parasitism, mutation accumulation and the maintenance of sex. *Nature* **367**, 554–557 (1994).
- Peters, A. D. & Lively, C. M. The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. *Am. Nat.* **154**, 393–405 (1999).
- Judson, O. P. Preserving genes: a model of the maintenance of genetic variation in a metapopulation under frequency-dependent selection. *Genetic Res.* **65**, 175–191 (1995).
- Gandon, S., Capowiez, Y., Dubois, Y., Michalakakis, Y. & Olivieri, I. Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proc. R. Soc. B* **263**, 1003–1009 (1996).

- Lively, C. M. Migration, virulence, and the geographic mosaic of adaptation by parasites. *Am. Nat.* **153**, S34–S47 (1999).
- Parker, M. Local population differentiation for compatibility in an annual legume and its host-specific fungal pathogen. *Evolution* **39**, 713–723 (1985).
- Lively, C. M. Adaptation by a parasitic trematode to local populations of its host. *Evolution* **46**, 1663–1671 (1989).
- Ebert, D. Virulence and local adaptation of a horizontally transmitted parasite. *Science* **256**, 1084–1086 (1994).
- Dybdahl, M. D. & Lively, C. M. Diverse, endemic and polyphyletic clones in mixed populations of the freshwater snail *Potamopyrgus antipodarum*. *J. Evol. Biol.* **8**, 385–398 (1995).
- Dybdahl, M. F. & Lively, C. M. Host–parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution* **52**, 1057–1066 (1998).
- Dybdahl, M. F. & Lively, C. M. The geography of coevolution: comparative population structures for a snail and its trematode parasite. *Evolution* **50**, 2264–2275 (1996).
- Norusis, M. J. *SPSS Advanced Statistics User's Guide* (SPSS, Chicago, 1990).

Acknowledgements

We thank L. Delp for helpful comments on the manuscript. This study was supported by the US National Science Foundation.

Correspondence and requests for materials should be addressed to M.F.D. (e-mail: dybdahl@ohiou.edu) or C.M.L. (e-mail: clively@indiana.edu).

Adhesive force of a single gecko foot-hair

Kellar Autumn*, Yiching A. Liang†, S. Tonia Hsieh‡, Wolfgang Zesch§, Wai Pang Chan‡, Thomas W. Kenny†, Ronald Fearing§ & Robert J. Full‡

* Department of Biology, Lewis and Clark College, Portland, Oregon 97219, USA

† Department of Mechanical Engineering, Stanford University, Stanford, California 94305, USA

‡ Department of Integrative Biology, University of California at Berkeley, Berkeley, California 94720, USA

§ Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, California 94720, USA

Geckos are exceptional in their ability to climb rapidly up smooth vertical surfaces^{1–3}. Microscopy has shown that a gecko's foot has nearly five hundred thousand keratinous hairs or setae. Each 30–130 μm long seta is only one-tenth the diameter of a human hair and contains hundreds of projections terminating in 0.2–0.5 μm spatula-shaped structures^{2,4}. After nearly a century of anatomical description^{2,4–6}, here we report the first direct measurements of single setal force by using a two-dimensional micro-electro-mechanical systems force sensor⁷ and a wire as a force gauge. Measurements revealed that a seta is ten times more effective at adhesion than predicted from maximal estimates on whole animals. Adhesive force values support the hypothesis that individual seta operate by van der Waals forces^{8,9}. The gecko's peculiar behaviour of toe uncurling and peeling² led us to discover two aspects of setal function which increase their effectiveness. A unique macroscopic orientation and preloading of the seta increased attachment force 600-fold above that of frictional measurements of the material. Suitably orientated setae reduced the forces necessary to peel the toe by simply detaching above a critical angle with the substratum.

The foot of a Tokay gecko (*Gekko gekko*) has about 5,000 setae mm^{-2} (ref. 4) and can produce 10 N of adhesive force with approximately 100 mm^2 of pad area¹⁰ (Fig. 1a–d). Therefore, each seta should produce an average force of 20 μN and an average stress of 0.1 N mm^{-2} (~ 1 atm). The actual magnitudes could be greater, as it is unlikely that all setae adhere simultaneously. We measured force production by single, isolated seta during attachment using a micromachined, dual-axis, piezoresistive cantilever (Fig. 1e).

To determine how setal force should be measured, we considered

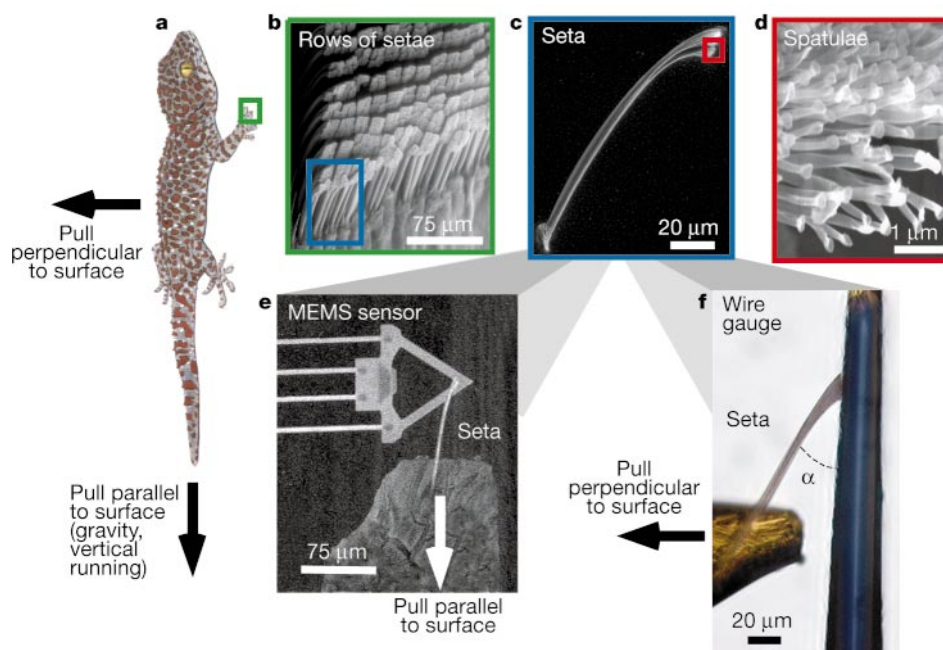


Figure 1 Gecko setae and apparatus for force measurement. **a**, Tokay gecko (*Gekko gecko*) with toe outlined. **b–d**, SEMs of rows of setae from a toe (**b**), a single seta (**c**) and the finest terminal branches of a seta, called spatulae (**d**). **e**, Single seta attached to a micro-electromechanical system (MEMS) cantilever⁷ capable of measuring force pro-

duction during attachment parallel and perpendicular to the surface. **f**, Single seta attached to an aluminum bonding wire capable of measuring force production during detachment perpendicular to the surface. Angle between setal stalk and wire represented by α .

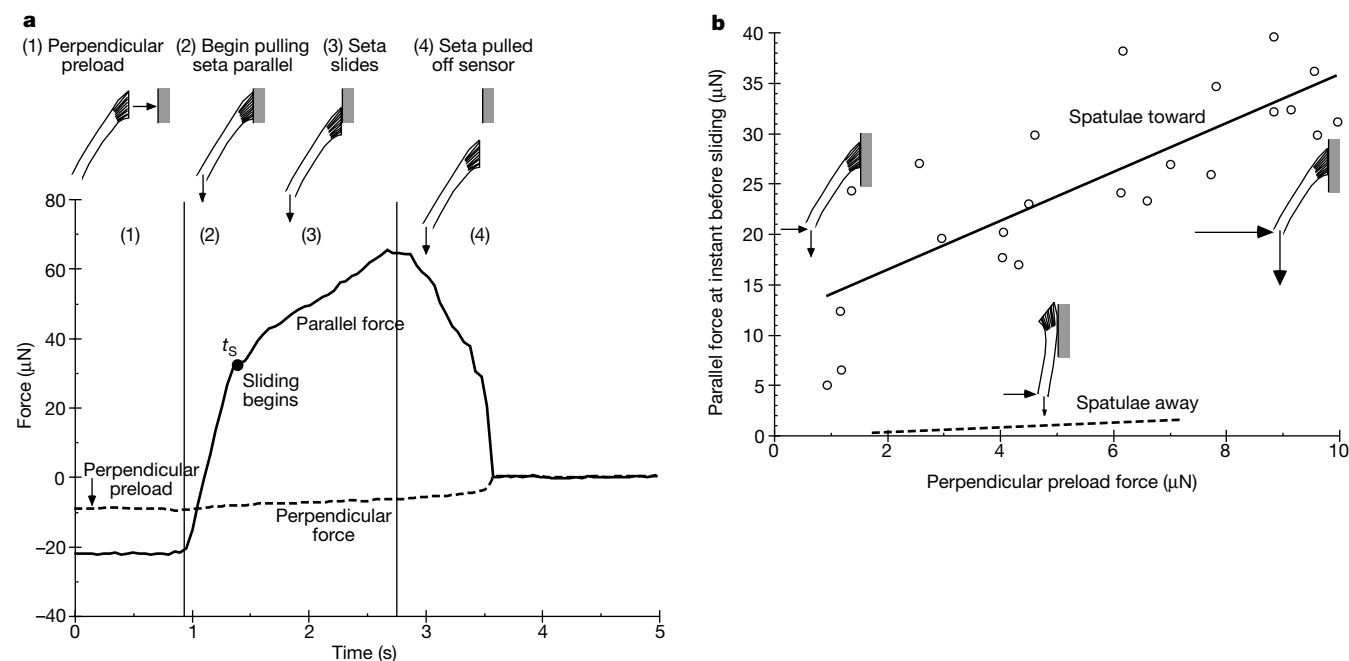


Figure 2 Force of single seta pulled parallel to the surface with a known perpendicular preload. **a**, Submaximal force (solid line) as a function of time. Perpendicular preload is designated by the dashed line. t_s represents the time when the seta began to slide off the sensor. The initial perpendicular force need not be maintained during the subsequent pull. Diagrams show the stages of setal movement corresponding to the force record. Arrows indicate the direction of applied force to the seta. Vertical arrow indicates a parallel force, and a horizontal arrow indicates a perpendicular force. Parallel force was zero prior to force application, and both parallel and perpendicular forces return to zero following force application. **b**, Setal force parallel to the surface during attachment as a function of perpendicular preload force. Setal force was taken to be the adhesive force at the time just prior to sliding (t_s). The solid line represents a seta with spatulae projecting toward the

surface. Results from a single seta are shown (parallel force = $2.8 \times$ perpendicular preload + 10.1; $r^2 = 0.74$; $n = 41$; $F = 113$; d.f. = 1, 39; $P < 0.0001$), but did not differ significantly in slope (analysis of covariance, variance ratio $F = 2.1$; degrees of freedom d.f. = 4, 57; $P = 0.10$) or intercept ($F = 0.052$; d.f. = 4, 57; $P = 0.99$) among five setae. The dashed line represents the setal force with spatulae projecting away from the surface (parallel force = $0.25 \times$ perpendicular preload - 0.09; $r^2 = 0.64$; $F = 13$; d.f. = 1, 9; $P = 0.007$). The force produced by the inactive, non-spatular region increased with normal or perpendicular force, typical of materials with a coefficient of friction equal to 0.25. The perpendicular preloading force that could be applied attained a maximum (near 15 μ N), because greater forces resulted in setal buckling.

the gecko's unusually complex behaviour of toe uncurling during attachment, which is much like blowing up an inflating party favour, and toe peeling during detachment, analogous to removing a piece of tape from a surface. The exquisite control of the toe allowed us to discover aspects of setal function by indicating that orientation and loading could be crucial to setal force capacity. Setal force did, in fact, depend on three-dimensional orientation (spatulae pointing towards or away from the surface) and the extent to which the seta was preloaded (pushed into and pulled along the surface) during the initial contact. Contacting the surface with the seta in a direction other than with spatulae projecting toward the surface resulted in forces of less than $0.3 \mu\text{N}$ when the seta was pulled away perpendicular to the surface. By contrast, when the active spatular region was projecting toward the surface, force increased enormously. After an initial push toward the surface, a 'perpendicular preload', the seta was pulled parallel to the surface. Setal adhesive force parallel to the surface increased until the seta began to slide off the edge of the sensor (at time t_s , Fig. 2a). Setal force parallel to the surface increased linearly with the perpendicular preloading force and was substantially greater than the force produced by the inactive, non-spatular region at all preloads (Fig. 2b). Experiments in which setae were pulled away from the surface of a wire (Fig. 1f) demonstrated that perpendicular preloading alone is insufficient for effective setal attachment. Setae that were first pushed into the surface and then pulled parallel to it developed over ten times the force ($13.6 \pm 2.6 \mu\text{N}$; $n = 17$) upon being pulled away from the surface than those having only a perpendicular preload ($0.6 \pm 0.7 \mu\text{N}$; $n = 17$). The largest parallel forces were observed only if the seta was allowed to slide approximately $5 \mu\text{m}$ along the sensor's surface, a distance imperceptible at the level of the foot (Fig. 3). The maximum adhesive force of single seta averaged $194 \pm 25 \mu\text{N}$ ($n = 28$), nearly tenfold greater than predicted from whole animal estimates. Our single-seta force measurements indicate that if all setae were simultaneously and maximally attached, a single foot of a gecko could produce 100 N of adhesive force ($\sim 10 \text{ atm}$). The results of preloading on setal force production support the hypothesis that a small perpendicular preloading force in concert with a rearward displacement or parallel preload may be necessary to engage adhesion⁴. Because the tips of the setae are directed rearwards away from the toenail, preloading may increase the number of spatulae contacting the surface¹¹.

The orientation of the setae was also important in detachment.

We found that setae detached at a similar angle ($30.6 \pm 1.8^\circ$; $n = 17$) when pulled away from the wire sensor's surface. To check for the presence of a critical angle of detachment, we controlled perpendicular force and progressively increased the setal angle (α ; Fig. 1f) until detachment. Setal angle at detachment changed by only 15% over a range of perpendicular forces (Fig. 4). This observation is consistent with an adhesive model where sliding stops when pulling at greater than the critical setal angle, and hence stress can increase at a boundary, causing fracture of the contact. Change in the orientation of the setae and perhaps even the geometry of the spatulae may help detachment. Geckos peel the tips of their toes away from a smooth surface during running². This toe peeling may have two effects. First, as we discovered here, it may put an individual seta in an orientation or at a critical angle that aids in its release. Second, toe peeling concentrates the detachment force on only a small subset of all attached setae at any instant.

Our direct setal force measurements reject two of the proposed mechanisms of adhesion, suction^{12,13} and friction^{6,14}, and provide indirect support for the most favoured hypothesis, intermolecular forces^{8,9}. Our measurements of greater than one atmosphere of adhesion pressure indicate that suction is not involved and support previous measurements carried out in a vacuum⁵. The present data do not support a friction mechanism^{6,14} because the coefficient of friction of the setal keratin on silicon is low ($\mu = 0.25$; Fig. 2b; dashed line). Microinterlocking¹⁴ could function as a secondary mechanism, but the cantilever's surface was smooth (surface roughness less than or equal to 2.5 nm) and the ability of geckos to adhere to polished glass shows that irregularities on the scale of the spatulae are not necessary for adhesion¹. Previous experiments using X-ray bombardment⁵ have eliminated electrostatic attraction¹⁵ as a mechanism necessary for setal adhesion, because the setae can still adhere in ionized air. Adhesion by glue is an unlikely mechanism, as skin glands are not present on the feet of lizards^{5,15,16}. However, the role of adsorbed water requires further study¹¹.

Our direct setal force measurements are consistent with the hypothesis that adhesion in geckos is the result of intermolecular forces^{8,9}. Earlier experimental support for the van der Waals hypothesis^{4,9} comes from the observation that adhesive force of a whole gecko increases with increasing surface energy of the substrate^{8,9}. The simple models available can only give the most approximate estimates of setal force production. If we assume that the tip of a spatula is a curved segment of a sphere (radius,

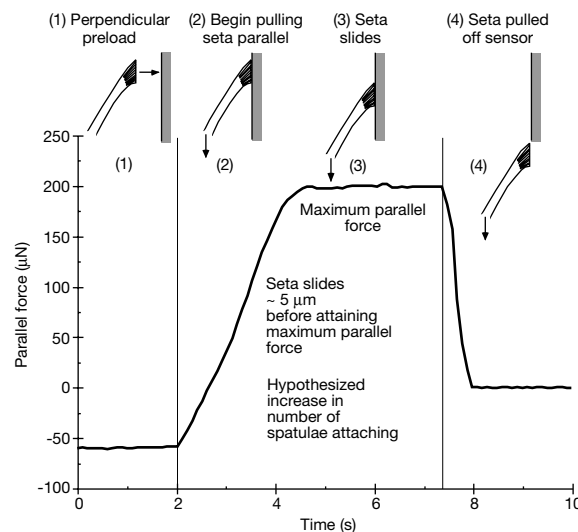


Figure 3 Maximal force after a maximum preload ($\sim 15 \mu\text{N}$) of a single seta parallel to the surface as a function of time. Diagrams show the stages of setal movement corresponding to the force record from the MEMS cantilever (Fig. 1e). Arrows indicate the direction of force applied to the seta. Vertical arrow indicates a parallel force; horizontal arrow

indicates a perpendicular force. The maximum force following the small rearward displacement ($\sim 5 \mu\text{m}$) was ten times that predicted from whole animal estimates (see text). The large increase in force during the rearward displacement may be caused by an increase in the number of spatulae contacting the surface.

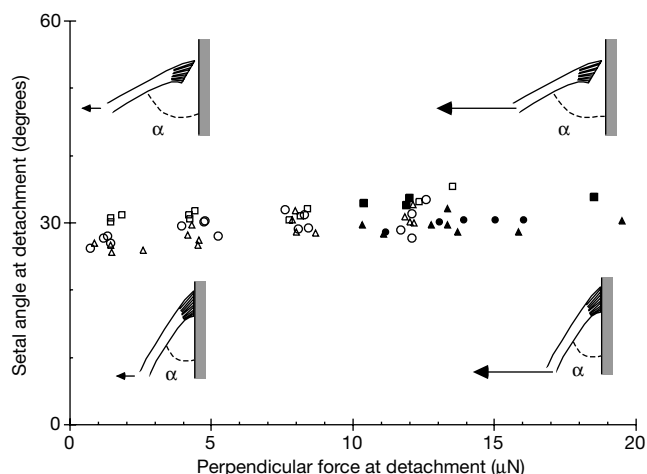


Figure 4 Setal angle (α) with the surface at detachment as a function of perpendicular force. Filled symbols represent seta pulled away from the surface until release. Open symbols represent seta held at a constant force as angle is increased. Each symbol shape represents a different seta. Data collected with wire gauge (Fig. 1f). Perpendicular force had a weak but significant effect on angle at detachment ($\alpha = 0.22 \times \text{perpendicular force} + 28.2$; $r^2 = 0.25$; $F = 20$; d.f. = 1, 59; $P < 0.0001$).

$R = 2 \mu\text{m}$) and is separated by a small distance from a large, flat surface where van der Waals forces become significant (atomic gap distance, $D \approx 0.3 \text{ nm}$), then setal force = $AR/6D^2$ where A is the material-dependent Hamaker constant taken to be 10^{-19} J (ref. 17). This estimate puts the van der Waals force for a spatula to be about $0.4 \mu\text{N}$. As the number of spatulae per seta varies from 100 to 1,000, setal force estimates range from 40 to $400 \mu\text{N}$, a range within which our direct measurements fall. The uncertainty in this estimate points to the need for future data collection on spatular morphology, orientation, spacing and material properties. van der Waals forces are extremely weak at greater than atomic distance gaps, and require intimate contact between the adhesive and the surface. Polymeric adhesives such as tape are soft, and are able to deform sufficiently for intimate contact over a relatively large surface area^{18,19}. The feet of *G. gecko* contain approximately one billion spatulae that could provide a sufficiently large surface area in close contact with the substrate⁴ for adhesion to be the result of van der Waals forces. Although manufacturing small, closely packed arrays mimicking setae are beyond the limits of human technology, the natural technology of gecko foot-hairs can provide biological inspiration for future design of a remarkably effective adhesive. □

Methods

Preparation of single seta

We carefully peeled the cuticular layer of a single row of lamellae off the toe of a restrained, live, non-moulting gecko. With a finely etched tungsten pin, we scraped the cuticular surface to break off individual setae at the base of the stalk. The isolated seta was then glued to the end of a #2 insect pin with 5-min epoxy (TTWDevcon, Danvers, Massachusetts). The pin had a tip diameter of approximately $15 \mu\text{m}$. To prevent the epoxy from creeping up the stalk of the seta, which might change the mechanical property of the specimen, we pre-cured the epoxy for about 1 min before applying it to the specimen. All setae were orientated so that the active surface was roughly perpendicular to the axis of the pin. All preparations were completed under a compound microscope.

Force estimation of a single seta during a parallel pull

To measure force parallel and perpendicular to the surface, we used a micromachined, dual-axis piezoresistive cantilever⁷ fabricated on a single-crystalline silicon wafer (Fig. 1e). It had two independent force sensors, each with one predominant direction of compliance. The perpendicular force sensor consisted of a thin triangular probe. The parallel force sensor was composed of four long slender ribs. A special 45° oblique ion implantation allowed piezoresistive and conductive regions to be implanted on both the parallel and perpendicular surfaces simultaneously. Forces applied to the tip of the sensor were resolved into these two orthogonal directions (parallel and perpendicular), and were measured by the changes in resistance of the piezoresistors. The minimum detectable force

for these cantilevers, calculated from the noise spectra, is $\sim 5 \text{ nN}$ in a 10 kHz bandwidth. Maximum force measurements possible exceed $300 \mu\text{N}$. The spring constant of the sensor was calibrated using a commercial force calibration cantilever (ThermoMicroscopes). The displacement sensitivity was obtained by measuring the resistance change of the piezoresistors while deflecting the cantilever by a known distance. As this device was originally designed for atomic force microscope data storage applications, each of these cantilevers had a sharp tip near the vertex of its triangular probe. For the gecko setae adhesion measurement, the back-side of this device was used to provide a smooth surface for setal adhesion.

Each seta was brought in contact with the sensor by applying a small preload perpendicular to the surface to increase contact and induce adhesion. To determine the effect of preload force on submaximal parallel force, we varied preload force when setae were attached to the tip of the sensor (Fig. 2b). To measure maximal parallel force, we used the base of the triangular probe. Using the base increased the area of contact, but did not allow for simultaneous measurement of preload forces. Sensor signals were taken while the seta was being pulled parallel to the surface by a piezoelectric manipulator at a rate of $\sim 5 \mu\text{m sec}^{-1}$. Sensor signals were amplified and filtered through a 300-Hz low-pass filter, and then digitized at 100 Hz using a 16-bit data acquisition card (National Instruments). The collected data (in volts) were converted to deflections of the sensor through the displacement sensitivity, and multiplied by the spring constant to obtain force values.

Force estimation of a single seta during a perpendicular pull

Breaking or detachment force was defined as the maximal force a seta could exert perpendicular to a surface immediately before it released. We determined this value for individual seta by measuring the amount it could displace a force gauge made from a 4.7 mm aluminum bonding wire with $25 \mu\text{m}$ nominal diameter (American Fine Wire Corp., Selman, Alabama; Fig. 1f). To maximize contact area of the active surface of the seta to the wire, we flattened a $50 \times 100 \mu\text{m}$ section at the wire's distal tip. The proximal end of the wire was fixed with epoxy onto a brass stub. We pressed the active surface of the seta against the flattened wire, producing a known perpendicular preload ($1.6 \pm 0.25 \mu\text{N SD}$). We applied a perpendicular detachment force to the seta using two different methods.

(1) We pulled the seta perpendicular to the wire. (2) We displaced the insect pin $19.7 \pm 3.45 \mu\text{m}$ along the wire to produce an additional parallel preload on the seta before pulling perpendicular to the wire. We then applied perpendicular detachment forces ranging from 0.5 to $20 \mu\text{N}$ and increased the angle of the seta with respect to the wire (α) until detachment occurred.

In all trials, detachment force was calculated from the maximum displacement of the wire pulled by the seta. All sequences were recorded with a video camera (Sony CCD) and digitized to a computer using a video editing system (Media 100 Inc., Marlboro, Massachusetts). The initial position of the wire, the angle of the seta with respect to the wire (α) and the position of the wire at the point of separation were recorded and analysed using image analysis software (NIH-Image). The amount of deflection in the force gauge was converted to adhesion force after we calibrated the force gauge against standard weights.

Received 31 January; accepted 27 March 2000.

- Maderson, P. F. A. Keratinized epidermal derivatives as an aid to climbing in gekkonid lizards. *Nature* **203**, 780–781 (1964).
- Russell, A. P. A contribution to the functional morphology of the foot of the tokay, *Gekko gecko* (Reptilia, Gekkonidae). *J. Zool. Lond.* **176**, 437–476 (1975).
- Cartmill, M. In *Functional Vertebrate Morphology* (eds Hildebrandt, M., Bramble, D. M., Liem, K. F. & Wake, D. B.) 73–88 (Harvard Univ. Press, Cambridge, Massachusetts, 1985).
- Ruibal, R. & Ernst, V. The structure of the digital setae of lizards. *J. Morphol.* **117**, 271–294 (1965).
- Dellit, W.-D. Zur anatomie und physiologie der Gekkozehe. *Jena Z. Naturw.* **68**, 613–656 (1934).
- Hora, S. L. The adhesive apparatus on the toes of certain geckos and tree frogs. *J. Proc. Asiatic Society of Bengal* **9**, 137–145 (1923).
- Chui, B. W., Kenny, T. W., Mamin, H. J., Terris, B. D. & Rugar, D. Independent detection of vertical and lateral forces with a sidewall-implanted dual-axis piezoresistive cantilever. *Appl. Phys. Lett.* **72**, 1388–1390 (1998).
- Hiller, U. Form und funktion der hautsinnesorgane bei gekkoniden. *Form. Funct.* **4**, 240–253 (1971).
- Hiller, U. Untersuchungen zum Feinbau und zur Funktion der Haftborsten von Reptilien. *Z. Morphol. Tiere* **62**, 307–362 (1969).
- Irschick, D. J. et al. A comparative analysis of clinging ability among pad-bearing lizards. *Biol. J. Linnaea Soc.* **59**, 21–35 (1996).
- Stork, N. E. Experimental analysis of adhesion of *Chrysolina polita* (Chrysomelidae: Coleoptera) on a variety of surfaces. *J. Exp. Biol.* **88**, 91–107 (1980).
- Gadow, H. *The Cambridge Natural History Vol. 8 Amphibia and Reptiles* (McMillan, London, 1901).
- Gennaro, J. G. J. The gecko grip. *Natural History* **78**, 36–43 (1969).
- Mahendra, B. C. Contributions to the bionomics, anatomy, reproduction and development of the Indian house gecko *Hemidactylus flaviviridis* Ruppell. Part II. The problem of locomotion. *Proc. Indian Acad. Sci.* **3**, 288–306 (1941).
- Schmidt, H. R. Zur Anatomie und Physiologie der Geckopfote. *Jena Z. Naturw.* **39**, 551 (1904).
- Bellairs, A. *The life of reptiles* (Universe, New York, 1970).
- Israelachvili, J. *Intermolecular & Surface Forces* (Academic, New York, 1992).
- Kinloch, A. J. *Adhesion and adhesives: science and technology* (Chapman and Hall, New York, 1987).
- Gay, C. & Leibler, L. Theory of tackiness. *Phys. Rev. Lett.* **82**, 936–939 (1999).

Acknowledgements

Supported by IS Robotics, a DARPA/ONR grant to R.J.F., an ONR MURI, NSF Graduate Research Fellowship, and NSF career and XYZ on a Chip awards to T.W.K. Fabrication of the MEMS devices used the Stanford Nanofabrication Facility. We thank S. Block,

K. Meijer, D. Dudek, A. Ahn, T. Kubow and J. Hearst for comments on earlier drafts, E. Florance for electron microscopy, the University of California Museum of Paleontology electron microscope laboratory and the Scientific Visualization Center at Berkeley.

Correspondence and requests for materials should be addressed to R.J.F. (e-mail: rjfull@socrates.berkeley.edu).

Neural synchrony correlates with surface segregation rules

Miguel Castelo-Branco, Rainer Goebel, Sergio Neuenschwander & Wolf Singer

Max-Planck-Institut für Hirnforschung, Deutschordenstraße 46, 60528 - Frankfurt am Main, Germany

To analyse an image, the visual system must decompose the scene into its relevant parts. Identifying distinct surfaces is a basic operation in such analysis, and is believed to precede object recognition^{1,2}. Two superimposed gratings moving in different directions (plaid stimuli) may be perceived either as two surfaces, one being transparent and sliding on top of the other (component motion) or as a single pattern whose direction of motion is intermediate to the component vectors (pattern motion)^{3–6}. The degree of transparency, and hence the perception, can be manipulated by changing only the luminance of the grating intersections^{7–12}. Here we show that neurons in two visual cortical areas—A18 and PMLS—synchronize their discharges when responding to contours of the same surface but not when responding to contours belonging to different surfaces. The amplitudes of responses correspond to previously described rate predictions^{3,13–16} for component and pattern motion, but, in contrast to synchrony, failed to reflect the transition from component to pattern motion induced by manipulating the degree of transparency. Thus, dynamic changes in synchronization could encode, in a context-dependent way, relations among simultaneous responses to spatially superimposed contours and thereby bias their association with distinct surfaces.

Specialized visual neurons may signal the motion of either the individual gratings of a plaid (component-selective cells) or of the global pattern (pattern-selective cells). For the identification of these cells, responses evoked by single gratings were compared to those evoked by unambiguous plaid patterns moving in an intermediate direction, and the results indicate that both cell types exist^{3,13–16}. However, except for a single study in the motion sensitive area (MT) of monkey visual cortex¹¹, no attempt has been made to investigate changes in response amplitudes associated with gradual modifications of transparency conditions. Therefore, little is known about how well individual cells differentiate between component and pattern motion.

Here, we investigate this question with multi-electrode recordings from neurons located in areas A18 and PMLS (postero-medial bank of the lateral suprasylvian sulcus) of the cat visual cortex. In addition, we use cross-correlation analysis to examine the hypothesis that binding of the moving contours into one coherent pattern or two independently moving gratings is associated with changes in the synchronization of responses.

Neurons synchronize their discharges with a precision in the millisecond range if they are activated by single contours, but they do not synchronize when activated by contours of different objects^{17–20}. It has been proposed, therefore, that synchronization serves as a binding mechanism by virtue of selectively raising the saliency of the synchronized discharges and thereby favouring their

joint processing at subsequent levels. Accordingly, when exposed to plaid stimuli, cells responding to contours of the same surface should synchronize their responses, and cells activated by contours belonging to different surfaces should not synchronize. Which cells are associated with a particular surface depends on transparency conditions, on the match of the cells' preferred direction of motion with the direction of motion of the plaid components and on the spatial relation (overlap, colinearity) of the receptive fields (RFs). In pattern motion, all cells capable of responding to the respective component motions should synchronize their responses because they are excited by contours of the same surface. In component motion, only those cells that respond to the contours of the same component grating should synchronize. This synchronization

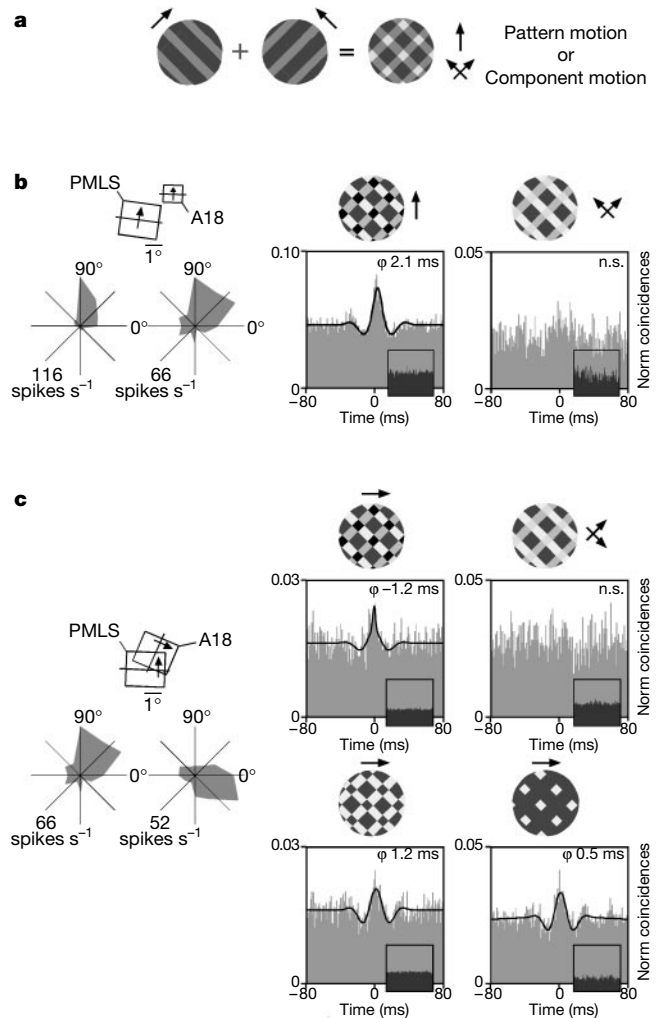


Figure 1 Dependence of synchrony on transparency conditions and receptive field (RF) configuration. **a**, Stimulus configuration. **b**, Synchronization between neurons with non-overlapping RFs and similar directional preferences recorded from A18 and PMLS. Left, RF constellation and tuning curves; right, cross-correlograms for responses to a non-transparent (left) and transparent plaid (right) moving in the cells' preferred direction. Grating luminance was asymmetric to enhance perceptual transparency². Small dark correlograms are shift predictors. **c**, Synchronization between neurons with different direction preferences recorded from A18 (polar and RF plots, left). Top, correlograms of responses evoked by a non-transparent (left) and a transparent (right) plaid moving in a direction intermediate to the cells' preferences. Bottom, correlograms of responses evoked by a non-transparent plaid with reversed contrast conditions (left), and by a surface defined by coherent motion of intersections (right). Scale on polar plots: discharge rate in spikes per second. Scale on correlograms: abscissa, shift interval in ms, bin width 1 ms; ordinate, number of coincidences per trial, normalized. Thick line, fitted gabor function; ϕ , phase shift. n.s., not significant.

should be particularly robust for cells with colinearly aligned, overlapping RFs as these respond to the same bar of a component grating.

Results of experiments addressing these predictions for cell pairs with different RF configurations are shown in Figs 1–3. Population data are presented in Fig. 4. Figure 1b shows neurons with non-overlapping and non-colinear RFs that have similar direction preference and were recorded simultaneously from areas A18 and PMLS. These cells synchronized in response to non-transparent patterns moving in the neurons' preferred direction but not when activated by a transparent plaid. Figure 1c (top) shows A18 neurons with non-overlapping RFs and different direction preferences. These cells too synchronized in the non-transparent condition when the direction of pattern motion was intermediate between the cells' preferred directions but not in the transparent condition. The synchronization in the non-transparent condition was of similar magnitude for all contrast conditions that bias perception towards pattern motion (Fig. 1c, compare top and bottom).

Figure 2 shows cells recorded from PMLS that had similar direction preferences and colinearly aligned overlapping RFs. These cells synchronized when activated with a single grating moving in their preferred direction and they continued to discharge in synchrony when another grating of different orientation and direction of motion was superimposed, even when transparency conditions were optimal for component motion. This indicates, as predicted, that the responses at both sites had remained associated with contours of the same surface.

To determine whether the degree of synchrony observed with plaids can be accounted for by linear summation of the synchrony induced by each grating alone, we calculated a linear predictor from the sum of coincidences evoked by each grating alone. For neuron pairs with colinear, overlapping RFs (Fig. 2a), synchrony to plaids did indeed correspond to the linear predictor. This was not so for cells with more discordant RFs and differing direction preferences (Fig. 2b). Their responses exhibited strong synchrony when evoked with one of the two gratings but this synchrony broke down when a second transparent grating was superimposed, indicating that addition of the second grating caused active desynchronization.

As predicted, cells at the two sites now respond independently to contours of different surfaces.

To exclude the possibility that the changes in synchronization induced by modifications of transparency were due to local contrast changes at the grating intersections rather than to global segmentation processes, we induced the transition from pattern to component motion by changing the ratio (duty cycle) between the width of white and dark bars of the gratings^{1,9} (Fig. 3, top). In this case, a gradual change in stimulus parameters leads to an abrupt change in figure–ground assignment. Non-transparent grating intersections suddenly change to background (figure–ground reversal) and this is associated with a sudden switch in perception from pattern to component motion^{1,9}. These modifications in duty cycle induce exactly the same changes in synchrony as modifications of transparency conditions by contrast manipulation of intersections (Fig. 3). The relations between synchronization strength, the transparency condition, the direction of plaid motion and the direction preferences of the neuron pairs conform to the predictions (Fig. 3b and c). Cells with differing direction preferences synchronize maximally with non-transparent patterns moving in directions intermediate to the respective preferred directions, but synchrony breaks down when the plaid becomes transparent (Fig. 3b). Some residual synchrony remains when the plaid moves in the opposite direction. In this case synchrony changes only little with transparency, indicating that now both sites respond mainly to contours of one grating. By contrast, synchrony among cells with overlapping, colinearly aligned RFs remains unaffected by changes in transparency. Synchrony varies with the direction of plaid movement but similarly for transparent and non-transparent plaids (Fig. 3c). This agrees with prediction as responses of cells with colinear RFs always remain associated with a single contour that may be either part of a pattern, or one of the two component gratings.

The perceptual switch between component and pattern motion is abrupt. To determine whether this is also the case for synchronization we measured correlation strength with a sliding window analysis while introducing gradual changes of duty cycle. For cells with different direction preference, synchronization broke down abruptly when the pattern became transparent. Correlation peaks

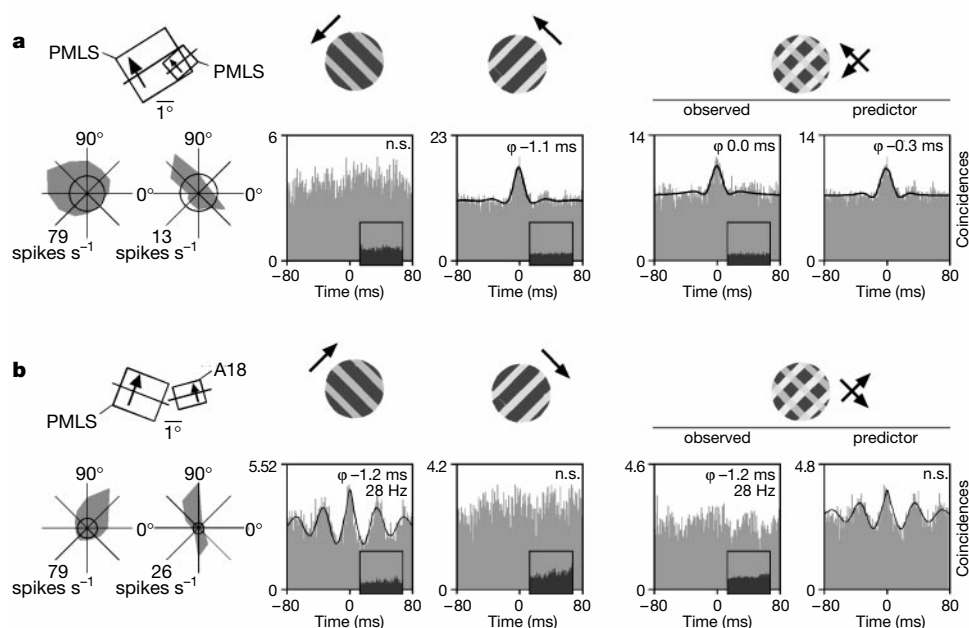


Figure 2 Comparison between single gratings and plaids. **a**, Synchronization between neurons with similar direction preferences and colinearly arranged, spatially contiguous RFs recorded from area PMLS. Correlograms are from responses to single gratings moving in the cells' non-preferred (first) and preferred direction (second) and from

responses to a transparent plaid resulting from a superposition of the two gratings (third correlogram). The fourth correlogram corresponds to the linear predictor. **b**, Same analysis as in **a** for neurons with spatially segregated RFs and differing direction preferences recorded from areas A18 and PMLS. Conventions as in Fig. 1.

remained high and unchanged during the first half of the duty cycle change and then dropped abruptly below the level of significance. By contrast, responses of cells with colinearly aligned, overlapping RFs remained synchronized throughout the change in duty cycle (Fig. 3d).

Classifying cells as component or pattern selective (see Methods and refs 13–16) confirmed that most cells in areas A18 and PMLS of cat visual cortex are best described as component selective (Fig. 4 top and refs 14–16). Furthermore, classification of cells into component- and pattern-selective subsets changed only little when response profiles were obtained with transparent and non-transparent plaids, respectively, indicating that individual cells do not differentiate between transparent and non-transparent conditions (Fig. 4a). To quantify this result, we subtracted the responses obtained with the two different plaid constellations from each other (component minus pattern condition) and reclassified the cells (Fig. 4, top right). Almost none of the resulting data points had any significant correlation with either of the two predictors. This proves that there is no systematic relation between rate changes of neurons in A18 and PMLS and transparency modifications.

Thus, differential synchronization might be the signal that biases perception towards component or pattern motion. To quantify stimulus-dependent changes in synchrony, we performed a split-plot analysis of variance (ANOVA). Synchronization strength was evaluated for each cell pair for responses evoked with single gratings and for responses evoked with both transparent and non-transparent plaids moving in the direction intermediate to those preferred by the respective cells. In the analysis, we also included a linear predictor of synchrony that corresponds to the correlation strength that one expects for plaids if correlations result simply from the sum of the coincidences in the responses to the respective single gratings. Correlation strength determined for the various conditions (pattern, component, single grating and linear predictor) was then related in a multifactorial analysis to the RF configuration of the analysed cell pairs. The following parameters were considered: RF overlap, colinearity of fields with similar direction preferences and difference between direction preferences. This analysis showed no detectable difference between cell pairs distributed within or across areas A18 and PMLS, so we pooled the data from all cell pairs. As predicted, responses of neuron pairs whose RFs differed in all the chosen categories (position, colinearity and direction preference) exhibit a highly significant reduction in synchrony when plaids undergo modifications that bias perception from pattern to component motion ($P \ll 0.01$, split-plot ANOVA, Fig. 4b). Conversely, no significant changes in synchrony ($P > 0.3$) occurred for cell pairs whose RFs had the same direction preference and were in addition either overlapping or colinear, conditions in which responses are likely to result from the same contour.

To illustrate the dependence of synchrony on both the transparency condition and the direction of plaid motion, we plotted synchronization strength against direction of plaid motion for transparent and non-transparent plaids as well as for a single grating (Fig. 4c). In this analysis we included only cells with strong direction preferences ($n = 25$) but pooled all pairs with similar and differing direction preferences. Differences in synchronization were maximal when plaid motion was in the direction intermediate to those preferred by the respective cells, the 'zero' direction. In this case non-transparent plaids induced strong synchronization whereas transparent plaids caused a complete break down of synchrony. When the direction of movement was offset from 'zero' by more than 30° , when one of the component gratings moved in a direction close to 'zero', differences between transparent and non-transparent conditions were less pronounced, because cells whose direction preference differed by less than 20° synchronized in both conditions. Single gratings induced synchrony at all directions of motion that evoked strong responses at both sites. This indicates that transparent plaids moving in 'zero' direction induce active desynchronization of

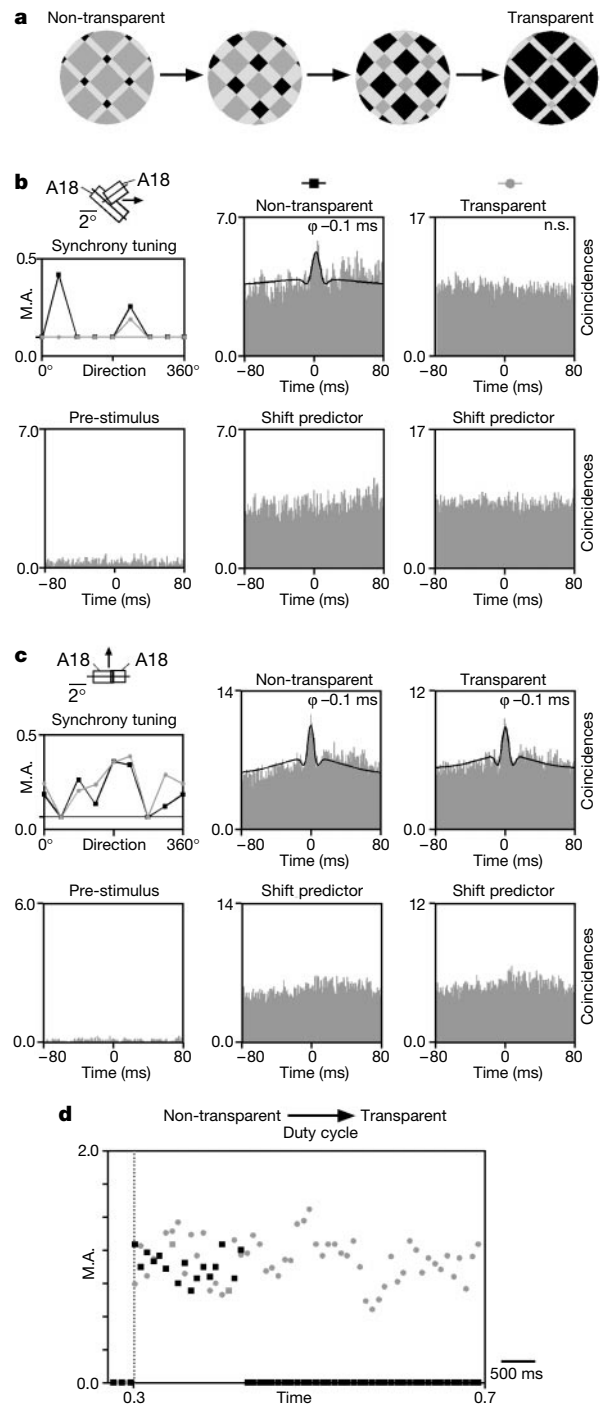


Figure 3 Changes in synchrony induced by manipulating the duty cycle of plaid components. **a**, A progressive change in duty cycle induces a change from non-transparent to transparent motion because of a flip in foreground–background assignment^{1,9}. **b**, **c**, Dependence of synchrony on the direction of plaid motion and transparency conditions for two cell pairs in A18 with discordant (**b**) and concordant (**c**) RFs. Left (synchrony tuning), relative modulation amplitude (MA) of correlation peaks for different directions of plaid motion (abscissa: 0° , horizontal right; 90° , vertical up) for non-transparent (dark squares) and transparent (grey circles) plaids. Middle and right, cross-correlograms for responses to non-transparent and transparent plaids moving in a direction intermediate between the cells' preferences (**b**) and in the cells' preferred direction (**c**). Bottom, pre-stimulus correlograms and shift predictors. **d**, Sliding window analysis (step 100 ms, window 400 ms) of changes of correlation strength (ordinate) resulting from duty cycle modifications (between 0.7 and 0.3, abscissa) for the cell pairs shown in **b** (black squares) and **c** (grey circles).

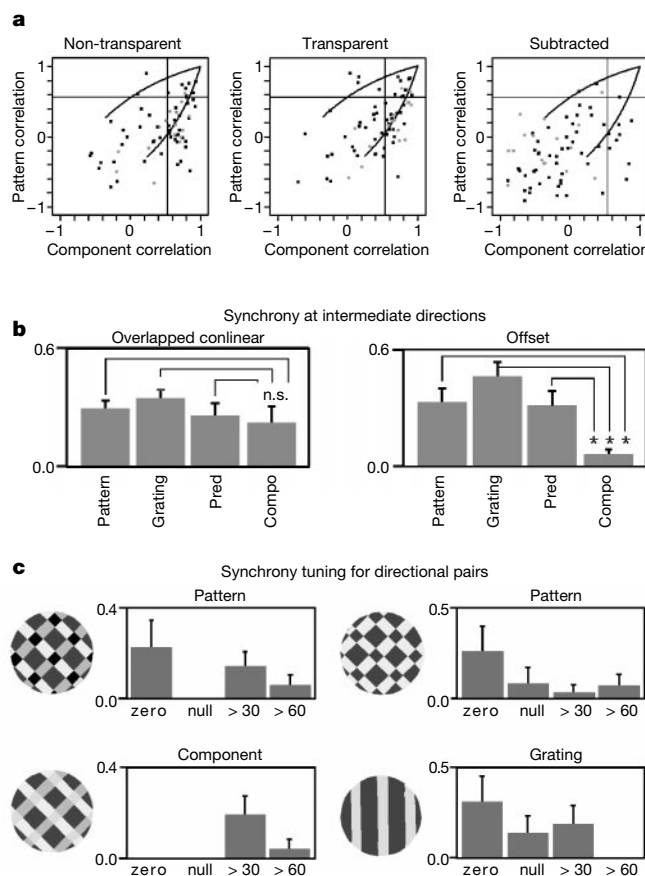


Figure 4 Effect of transparency manipulation on response amplitudes and synchrony. **a**, Scatter plots identifying cells as component (abscissa) or pattern selective (ordinate) obtained from responses to non-transparent (left) and transparent plaids (middle). Continuous lines indicate significance of classification. Cells in upper left and lower right cases are classified as pattern and component selective, respectively. Cells outside these regions or inside the curved area are unclassified. Right, classification after subtraction of responses to non-transparent plaids from those to transparent plaids. Dark squares and grey circles, stimuli in which either the duty cycle (squares) or the luminance of intersections (circles) have been manipulated. **b**, Comparison of correlation strength

(ordinate) for non-transparent (pattern) and transparent (component) plaids, single gratings (grating) and the linear predictor (pred) for cell pairs with overlapping, colinear (top) and non-overlapping, dissimilar RFs (bottom). Responses were evoked with plaids (gratings) moving in a direction intermediate to those preferred by the respective cells. **c**, Dependence of correlation strength (ordinate) on stimulus conditions (pattern, component, grating) and direction of stimulus motion (abscissa) for the subset of cells with strong directional preferences ($n = 25$). Zero, stimuli moving in a direction intermediate ($\pm 30^\circ$) to the cells' preferred directions; null, direction opposite to zero; > 30 and > 60 , directions offset by at least 30° or 60° from zero.

cells responding to the respective components.

In early visual areas such as those investigated here, most cells are component specific and, as our data indicate, their rate-modulated responses do not distinguish between pattern and component motion conditions^{13–16}. This calls for a transparency-sensitive, rate-independent grouping cue that permits subsequent processing stages to associate responses with either a single or two different surfaces. We propose that the differential synchronization of the responses of component-selective cells described here could serve as the required grouping signal, because synchrony occurred selectively among exactly those responses that need to be bound to account for the different perceptions induced by plaids of different transparency.

For the read-out of these differential synchronization patterns, two non-exclusive possibilities may be considered. Coincidence detectors in higher processing areas could transform the synchronized output of cells in A18 and PMLS in rate codes for the various surfaces in plaid stimuli. Alternatively, the synchronously active ensembles of neurons distributed across areas A18 and PMLS, and presumably also other cortical areas, could themselves serve as the distributed representations of the surfaces. Changes in cortical synchronization patterns are sensed by output structures such as the optic tectum²¹ and can thus lead to modification of eye movements without requiring further decoding²².

In conclusion, the present findings provide evidence that synchronization correlates well with the perceptual segmentation of a complex visual pattern into distinct, spatially overlapping surfaces. As no such correlation was found for the rate-modulated responses of neurons in the cortical areas analysed here, we propose that synchronization serves to group responses for further joint evaluation. □

Methods

Anaesthesia was induced with ketamine (Ketanest, Parke-Davis, 10 mg kg^{-1} , intramuscular) and xylazine (Rompun, Bayer, 2 mg kg^{-1} , intramuscular), and maintained with a mixture of 70% N_2O and 30% O_2 supplemented by 0.5–1.0% halothane. Paralysis was obtained with pancuronium bromide (Pancuronium, Organon, $0.15 \text{ mg kg}^{-1} \text{ h}^{-1}$). Single- or multi-unit activity (306 recording pairs in six animals) was recorded with multiple (up to four) varnish-coated tungsten electrodes²³. Cortical areas A18 and PMLS were identified on the basis of RF properties and histology²⁴. Signals were band-pass filtered from 1 to 3 kHz and spikes were detected with an amplitude discriminator with the threshold set to twice the noise level.

Visual stimuli were generated on a computer screen (refresh rate of 100 Hz). Squarewave gratings were superimposed in a circular display subtending $12\text{--}18^\circ$ of visual angle. Transparent and non-transparent plaids were interleaved with single gratings and moved in up to twelve different directions, movement direction changing from trial to trial in a pseudorandom sequence (20 trials per configuration).

Plaid velocities (pattern velocity, 8 and 16° s^{-1}), background, grating and intersection luminances ($0.5\text{--}140 \text{ cd m}^{-2}$, mean stimulus luminance $\sim 35 \text{ cd m}^{-2}$), duty cycle (around 0.3) and angle between gratings (preferably above 135°) were adjusted to induce strong perceptual biases. Stimulus conditions were validated by analysing eye movements from

electro-oculographic recordings in two awake cats and this indicated that the animals distinguished between component and pattern motion.

Auto- and cross-correlograms were calculated for responses from individual trials with a resolution of 1.0 ms and averaged over 20 stimulus presentations. Pre-stimulus correlograms controlled for stimulus independent correlations and shift predictors for synchronization induced by stimulus-locking²⁵. A damped cosine function was fitted to the correlograms to quantify the strength and phase shift of correlations and the frequency and modulation depth of oscillatory patterning^{23,26}. Correlation strength was assessed from the modulation amplitude (MA), the ratio between the amplitude of the central peak of both the fitted and the raw function (as a control) and the correlogram offset. MA values obtained with and without normalization to the geometric mean of firing rates (normalized coincidences) gave similar results.

A split-plot design²⁷ was applied (ANOVA for repeated measures) to evaluate correlation strength as a function of stimulus (pattern and component motion, single gratings, linear predictor derived from responses to single gratings) and RF configuration. Orientation and direction preferences, and tuning width were determined with vector averaging methods^{28,29}. Recording sites were classified as having similar orientation preferences if these differed by less than 20°, as having colinear RFs if the offset of the axial alignment of the RFs was < 20% of the size of the smaller RF, and as having overlapping RFs if discharge regions (mapped with small moving bars) overlapped. RFs were classified as directional if the ratio of responses to stimuli in preferred and non-preferred directions was ≥ 2 (true for 106 pairs). 55 pairs differed and 16 pairs matched in all categories; 25 pairs exhibited strong directional tuning at both sites.

Based on firing rate, neurons were classified as component selective if the directional tuning curve obtained with plaids correlated with that resulting from linear summation of the responses to the respective component gratings. Conversely, cells were classified as pattern-selective if the tuning curve obtained with a plaid correlated better with that resulting from stimulation with a single grating. Computation of statistical boundaries took into account the number of stimulus directions and the number of stimulus trials. The partial correlations were computed with the following equation:

$R_p = (r_p - r_{pc}r_{pc}) / [(1 - r_p^2)(1 - r_{pc}^2)]^{1/2}$. R_p , partial correlation coefficient for the pattern prediction; r_p , correlation coefficient between the plaid response and the component model; r_{pc} , correlation coefficient between the plaid response and the pattern model; r_{pc} , correlation coefficient for the two models. Also R_c is the partial correlation coefficient for the component prediction, obtained by interchanging r_p and r_c ¹³.

Received 24 January; accepted 11 April 2000.

1. Stoner, G. R. & Albright, T. D. The interpretation of visual motion: evidence for surface segmentation mechanisms. *Vision Res.* **36**, 1291–1310 (1996).
2. Shimojo, S., Silverman, G. H. & Nakayama, K. An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature* **333**, 265–268 (1988).
3. Movshon, J. A., Adelson, E. H., Gizzi, M. & Newsome, W. T. in *Study Group on Pattern Recognition Mechanisms* (eds Chagas, C., Gatass, R. & Gross, C. G.) (Pontificia Academia Scientiarum, Vatican City, 1985).
4. Adelson, E. H. & Movshon, J. A. Phenomenal coherence of moving visual patterns. *Nature* **300**, 523–525 (1982).
5. Stoner, G. R., Albright, T. D. & Ramachandran, V. S. Transparency and coherence in human motion perception. *Nature* **344**, 153–155 (1990).
6. Smith, A. T. & Harris, L. R. Use of plaid patterns to distinguish the corticofugal and direct retinal inputs to the brainstem optokinetic nystagmus generator. *Exp. Brain Res.* **86**, 324–332 (1991).
7. Metelli, F. The perception of transparency. *Sci. Am.* **230**, 90–98 (1974).
8. Beck, J. & Ivry, R. On the role of figural organization in perceptual transparency. *Percept. Psychophys.* **44**, 585–594 (1988).
9. Albright, T. D. & Stoner, G. R. Visual motion perception. *Proc. Natl Acad. Sci. USA* **92**, 2433–2440 (1995).
10. Stoner, G. R. & Albright, T. D. Luminance contrast affects motion coherence in plaid patterns by acting as depth from occlusion cue. *Vision Res.* **38**, 387–401 (1998).
11. Stoner, G. R. & Albright, T. D. Neural correlates of perceptual motion coherence. *Nature* **358**, 412–414 (1992).
12. Castelo-Branco, M. *et al.* MT/MST activation depends on the interpretation of stimuli: a functional fMRI study of the perception of plaids. *Soc. Neurosci. Abstr.* **23**, 460 (1997).
13. Rodman, H. R. & Albright, T. D. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp. Brain Res.* **75**, 53–64 (1989).
14. Gizzi, M. S., Katz, E., Schumer, R. A. & Movshon, J. A. Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *J. Neurophysiol.* **63**, 1529–1543 (1990).
15. Scannell, J. W. *et al.* Visual motion processing in the anterior ectosylvian sulcus of the cat. *J. Neurophysiol.* **76**, 895–907 (1996).
16. Merabet, L., Desautels, A., Minville, K. & Casanova, C. Motion integration in a thalamic visual nucleus. *Nature* **396**, 265–268 (1998).
17. Gray, C. M., Konig, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
18. Engel, A. K., Konig, P., Kreiter, A. K. & Singer, W. Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* **252**, 1177–1179 (1991).
19. Freiwald, W. A., Kreiter, A. K. & Singer, W. Stimulus dependent intercolumnar synchronization of single unit responses in cat area 17. *NeuroReport* **6**, 2348–2352 (1995).
20. Kreiter, A. K. & Singer, W. Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J. Neurosci.* **16**, 2381–2396 (1996).
21. Brecht, M., Singer, W. & Engel, A. K. Correlation analysis of corticotectal interactions in the cat visual system. *J. Neurophysiol.* **79**, 2394–2407 (1998).
22. Brecht, M., Singer, W. & Engel, A. K. Collicular saccade vectors defined by synchronization. *Soc. Neurosci. Abstr.* **23**, 843 (1997).
23. Castelo-Branco, M., Neuenschwander, S. & Singer, W. Synchronization of visual responses between

the cortex, lateral geniculate nucleus, and retina in the anesthetized cat. *J. Neurosci.* **18**, 6395–6410 (1998).

24. Palmer, L. A., Rosenquist, A. C. & Tusa, R. J. The retinotopic organization of lateral suprasylvian visual areas in the cat. *J. Comp. Neurol.* **177**, 237–256 (1978).
25. Perkel, D. H., Gerstein, G. L. & Moore, G. P. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* **7**, 419–440 (1967).
26. König, P. A method for the quantification of synchrony and oscillatory properties of neuronal activity. *J. Neurosci. Methods* **54**, 31–37 (1994).
27. Edwards, A. L. *Multiple Regression Analysis and the Analysis of Variance and Covariance* (W. H. Freeman, New York, 1985).
28. Batschelet, E. *Circular Statistics in Biology* (Academic, London, 1993).
29. Swindale, N. V. Orientation tuning curves: empirical description and estimation of parameters. *Biol. Cybern.* **78**, 45–56 (1998).

Acknowledgements

We thank H. Klon-Lipok, P. Janson and S. Grimm for their technical assistance in electrode manufacture and histology. This research was sponsored by the Max-Planck-Gesellschaft. M.C.-B. was partially supported by the Gulbenkian Foundation and Programa Praxis, Portugal.

Correspondence and requests for materials should be addressed to: W.S. (e-mail: singer@mpih-frankfurt.mpg.de).

Photoactivated γ -secretase inhibitors directed to the active site covalently label presenilin 1

Yue-Ming Li*, Min Xu*, Ming-Tain Lai*, Qian Huang*, José L. Castro†, Jillian DiMuzio-Mower†, Timothy Harrison†, Colin Lellis*, Alan Nadin†, Joseph G. Meduvelil†, R. Bruce Register*, Mohinder K. Sardana*, Mark S. Shearman†, Adrian L. Smith†, Xiao-Ping Shi*, Kuo-Chang Yin*, Jules A. Shafer* & Stephen J. Gardell*

* Department of Biological Chemistry, Merck Research Laboratories, West Point, Pennsylvania 19486, USA

† Neuroscience Research Centre, Merck Sharp & Dohme Research Laboratories, Terlings Park, Harlow, Essex CM20 2QR, UK

Cleavage of amyloid precursor protein (APP) by the β - and γ -secretases generates the amino and carboxy termini, respectively, of the A β amyloidogenic peptides A β 40 and A β 42—the major constituents of the amyloid plaques in the brain parenchyma of Alzheimer's disease patients¹. There is evidence that the polytopic membrane-spanning proteins, presenilin 1 and 2 (PS1 and PS2), are important determinants of γ -secretase activity: mutations in PS1 and PS2 that are associated with early-onset familial Alzheimer's disease^{2,3} increase the production of A β 42 (refs 4–6), the more amyloidogenic peptide; γ -secretase activity is reduced in neuronal cultures derived from PS1-deficient mouse embryos⁷; and directed mutagenesis of two conserved aspartates in transmembrane segments of PS1 inactivates the ability of γ -secretase to catalyse processing of APP within its transmembrane domain⁸. It is unknown, however, whether PS1 (which has little or no homology to any known aspartyl protease) is itself a transmembrane aspartyl protease or a γ -secretase cofactor, or helps to colocalize γ -secretase and APP. Here we report photoaffinity labelling of PS1 (and PS2) by potent γ -secretase inhibitors that were designed to function as transition state analogue inhibitors directed to the active site of an aspartyl protease. This observation indicates that PS1 (and PS2) may contain the active site of γ -secretase. Interestingly, the intact, single-chain form of wild-type PS1 is not labelled by an active-site-directed photoaffinity probe, suggesting that intact wild-type PS1 may be an aspartyl protease zymogen.

We used L-685,458 (Fig. 1), a potent γ -secretase inhibitor⁹, to prepare photoaffinity probes that label the active site of γ -secretase.

L-685,458 contains a hydroxyethylene dipeptide isostere that should serve as a transition state analogue to direct the inhibitor to the active site of an aspartyl protease target^{10,11}. The mode of action of L-685,458 was investigated by comparing its inhibitory potency with that of its hydroxy epimer, L-682,679 (Fig. 1), and corresponding peptide analogue, L-405,484 (Fig. 1). The assay consisted of a recombinant APP-related substrate, C100Flag (Fig. 2a), and solubilized γ -secretase from HeLa cell membranes, which constitutes virtually all of the cellular γ -secretase activity¹². Cleavage at the γ -secretase scissile bond in C100Flag was detected by electrochemiluminescence using an antibody that specifically recognizes the C terminus of A β 40. Inversion of the configuration at the hydroxylic carbon atom of the hydroxyethylene dipeptide isostere in L-695,458 (IC_{50} 0.3 nM) to yield L-682,679 (IC_{50} 82 nM) reduces inhibitory potency 270-fold (Fig. 2b). Moreover, L-405,484, the peptide analogue of L-685,458, only weakly inhibits the processing of C100Flag (IC_{50} > 100 μ M) by solubilized γ -secretase (Fig. 2b). The ability of L-405,484 to inhibit the hydrolysis of C100Flag catalysed by γ -secretase probably reflects its ability to serve as a competing γ -secretase substrate, as γ -secretase catalyses the release of *t*-butyloxycarbonyl-Phe from L-405,484 in a reaction that is inhibited by L-685,458 (data not shown). The observations that the inhibitory potency of L-685,458 is sensitive to the configuration of the hydroxylic carbon atom and the substrate analogue of L-685,458 is processed by γ -secretase and binds to the enzyme much less tightly than L-685,458 provide strong evidence that L-685,458 is an active-site-directed transition state analogue inhibitor for an aspartyl protease^{10,11}. Our conclusion that γ -secretase is an aspartyl

protease agrees with that of a previous study showing that other aspartyl protease transition state analogues inhibit γ -secretase activity in cells¹³.

We prepared the benzophenone-containing compounds L-852,505 and L-852,646 (Fig. 1), reasoning that photoreactive derivatives of L-685,458 might also be directed to the active site of the enzyme. The photoreactive benzophenone group was placed at the N terminus in L-852,505 and near the C terminus in L-852,646 so that the two photoaffinity probes might label distinct domains within the active site of γ -secretase. A biotin moiety was incorporated in the photoaffinity probes to facilitate isolation and identification of the adducts. The interaction of the photoaffinity probes with γ -secretase was studied using solubilized γ -secretase¹² to circumvent complications arising from limited uptake by cells of the biotinylated photoreactive compounds. In the absence of photoactivation, L-852,505 and L-852,646 display excellent inhibitory potency (IC_{50} $\approx$ 1 nM) towards solubilized γ -secretase (Fig. 2b). L-852,505 and L-852,646 were photoactivated in the presence of solubilized γ -secretase to covalently label the active site of their aspartyl protease targets. Following photoactivation, the reaction mixture was treated with RIPA buffer to increase the accessibility of the biotin group in the photolabelled proteins to streptavidin. The biotin-linked photolabelled proteins were isolated from the reaction mixture by absorption on streptavidin-agarose beads. The covalently labelled proteins were then eluted from the beads, subjected to SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and visualized by immunoblotting using an anti-biotin antibody. In the case of L-852,505, we see a major biotinylated species (P20) with an apparent relative molecular mass (M_r) of 20,000 (20K) (Fig. 3a). Also evident with photoactivated L-852,505 is a lesser biotinylated species (P23) with an M_r of 23K. Similar results are obtained when L-852,505 is photoactivated in the presence of intact HeLa cell membranes (data not shown). Photoactivation of L-852,646 in the presence of solubilized γ -secretase causes covalent labelling of a species (P34) with an apparent M_r of 34K (Fig. 3b). We obtained no covalently labelled biotinylated

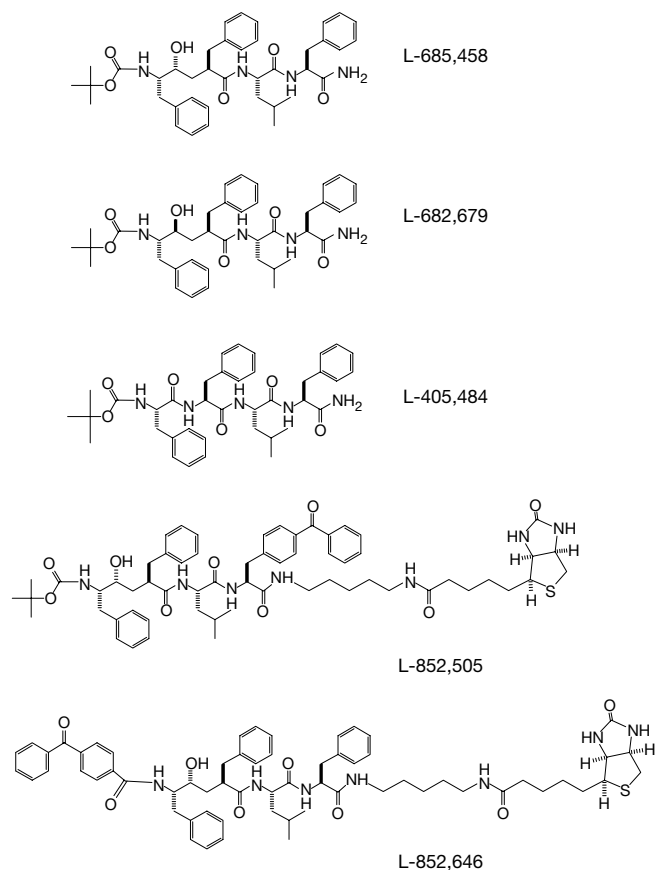


Figure 1 Chemical structures of the γ -secretase inhibitor, L-685,458 and its derivatives: L-682,679 (hydroxy group epimer), L-405,484 (peptidic analogue), L-852,505 and L-852,646 (both photoreactive/biotinylated analogues). Note that the photoreactive benzophenone moiety is situated at different positions in the two photoreactive derivatives of L-685,458.

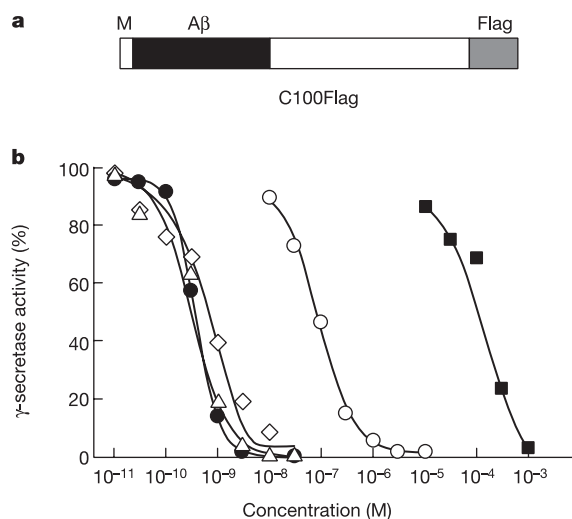


Figure 2 Inhibitory potencies of L-685,458 and its derivatives in the *in vitro* γ -secretase assay. **a**, Schematic representation of the fusion protein substrate (C100Flag) used for the *in vitro* γ -secretase assay. C100Flag consists sequentially of an N-terminal Met (M), APP597-695 (the A β domain is indicated) and the Flag tag (FLAG) sequence. **b**, Inhibition of *in vitro* γ -secretase activity. C100Flag was added to CHAPSO-solubilized HeLa cell membranes containing the indicated concentrations of L-685,458 (filled circle), L-682,679 (open circle), L-405,484 (filled square), L-852,505 (open triangle) and L-852,646 (open diamond). After 90 min, aliquots were assayed for A β (x-40) by electrochemiluminescence. Activity levels are relative to that observed in the absence of inhibitor (assigned 100%).

species when solubilized γ -secretase was treated with L-852,505 or L-852,646 without photoactivation (data not shown). The observation that the non-photoreactive γ -secretase inhibitor L-685,458 blocks covalent labelling of P20, P23 and P34 (Fig. 3) suggests that binding to each of these species is specific and not due to the presence of the benzophenone or biotin groups in the photoaffinity probes.

The growing evidence that γ -secretase and PS1 are closely related^{7,8,12} prompted us to explore whether it is actually PS1 that is photolabelled by L-852,505 and L-852,646. PS1 is proteolytically cleaved to yield a non-covalent heterodimer (Fig. 3c) involving an N-terminal fragment (PS1-NTF; $M_r \sim 30K$) and a C-terminal fragment (PS1-CTF; $M_r \sim 20K$)¹⁴. The biotinylated species in L-852,505- or L-852,646-treated solubilized γ -secretase were analysed after SDS-PAGE by immunoblotting using antibodies against PS1. Testing the biotinylated species in L-852,505-treated solubilized γ -secretase with an antibody against the C terminus of PS1-CTF shows that P20 and P23 are immunoreactive (Fig. 3a). The appearance of the immunoreactive PS1-CTF species is blocked when photoactivation of L-852,505 is performed in the presence of the non-photoreactive derivative L-685,458. The PS1-NTF fragment is not observed when L-852,505-treated, biotinylated proteins are probed with the anti-PS1-NTF antibody. Probing the biotinylated species in L-852,646-treated solubilized γ -secretase with an antibody against the PS1-NTF subunit shows that P34 is immunoreactive (Fig. 3b). As with L-852,505, formation of this immunoreactive species is blocked when photoactivation with L-852,646 is performed in the presence of L-685,458. Immunoreactive bands are not observed when the L-852,646-treated, biotinylated proteins are probed with an antibody against the PS1-CTF subunit. These observations indicate that photoactivated L-852,505 and L-852,646 covalently label PS1-CTF and PS1-NTF, respectively. The P23 species is probably a phosphorylated form of PS1-CTF¹⁵. Dissociation of the PS1 heterodimer by the detergent-containing RIPA buffer, as shown with Triton X-100¹⁶, probably accounts for our inability to capture and visualize both PS1-CTF and PS1-NTF when one subunit, but not both, is covalently labelled with a

biotinylated photoaffinity probe.

The lack of biotin immunoreactive bands other than P20, P23 or P34 (Fig. 3a, b) indicates that only PS1 or proteins that migrate with the PS1 fragments are labelled with the photoaffinity probes in the molecular mass range probed in the gels in Fig. 3. A search for biotin immunoreactive bands up to 250K revealed only one species ($M_r \sim 80K$, data not shown). This band, which was not dependent on photoactivation or blocked by L-685,458, is attributed to an endogenous biotinylated protein that was not removed by the pretreatment of solubilized γ -secretase with the streptavidin beads (see Methods).

Because of the homology between PS2 and PS1 (67% identity), we wished to assess whether PS2 was labelled by the photoaffinity agents. We used an antibody against the C terminus of PS2-CTF to show that photoactivation of L-852,505 in the presence of solubilized γ -secretase results in the covalent labelling of PS2-CTF (Fig. 3a). The covalent labelling of PS2-CTF was blocked when photoactivation was performed in the presence of the non-photoreactive derivative L-685,458. These observations suggest that PS2, like PS1, contains the active site of a γ -secretase. PS2 may indeed catalyse the residual γ -secretase activity in neuronal cultures derived from PS1-deficient mouse embryos⁷.

The low content of full-length PS1 in our solubilized γ -secretase preparation precluded assessment of whether intact PS1 binds the photoaffinity probes. However, heterologous expression of PS1 in mammalian cells leads to accumulation of unprocessed PS1 (ref. 14). We therefore prepared solubilized γ -secretase from murine N2a neuroblastoma cells that express human PS1 (N2a/PS1 solubilized γ -secretase). SDS-PAGE/immunoblot analysis of N2a/PS1 solubilized γ -secretase with antibodies against PS1-NTF and PS1-CTF confirmed the presence of the intact form of PS1 as well as the individual fragments of the PS1 heterodimer (Fig. 4a). Photoactivation of L-852,505 in the presence of N2a/PS1 solubilized γ -secretase yields a biotinylated protein that migrates with PS1-CTF and reacts with the anti-PS1-CTF antibody. The appearance of this biotinylated protein is blocked when the photoactivation is performed in the presence of excess L-685,458. Full-length PS1, which displays a

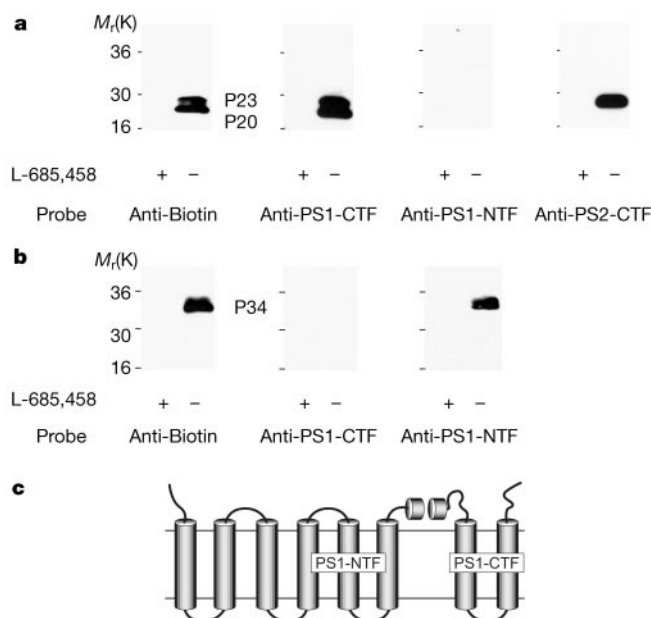


Figure 3 Covalent labelling of PS1 by photoreactive active site directed γ -secretase inhibitors. L-852,505 and L-852,646 (**a**, **b**, respectively) were photoactivated with CHAPSO-solubilized HeLa cell membranes in the absence or presence of L-685,458. The sample was diluted with RIPA buffer. Biotinylated proteins were captured with streptavidin-agarose and probed by SDS-PAGE/immunoblotting using anti-biotin

antibody against the C terminus of PS1-CTF (anti-PS1-CTF), an antibody against the N terminus of PS1-NTF (anti-PS1-NTF) or an antibody against the C terminus of PS2-CTF (anti-PS2-CTF). The mobilities of the P20, P23 and P34 species as well as molecular mass standards are shown. **c**, Schematic representation of the PS1 heterodimer comprising the PS1-NTF and PS1-CTF fragments.

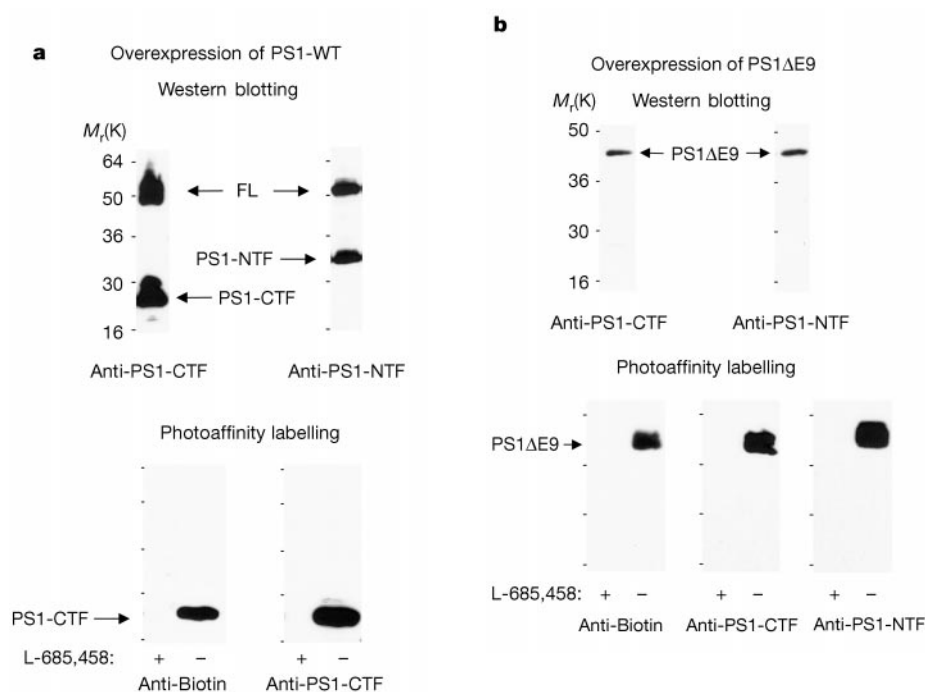


Figure 4 Photoactivated γ -secretase inhibitor does not bind covalently to intact wild-type PS1 but reacts with the PS1-FAD ('familial Alzheimer's disease') Δ exon 9 mutant. Membrane extracts were prepared from N2a neuroblastoma cells stably transfected with PS1 (**a**) and N2a neuroblastoma cells stably transfected with the PS1-FAD Δ exon 9 mutant (**b**). The membrane extracts were analysed by SDS-PAGE/immunoblotting using anti-PS1-NTF and anti-PS1-CTF antibodies. L-852,505 was added to the membrane

extracts (in the absence and presence of excess L-685,458) and subjected to photoactivation. The ensuing biotinylated proteins were diluted with RIPA buffer, captured with immobilized streptavidin and analysed by SDS-PAGE/immunoblotting using anti-biotin, anti-PS1-CTF and anti-PS1-NTF antibodies. The mobilities of intact wild-type PS1 (FL), PS1-NTF, PS1-CTF and PS1-FAD Δ exon 9 (PS1 Δ E9) as well as molecular mass standards are shown.

robust signal in the immunoblot of N2a/PS1 solubilized γ -secretase, is not biotinylated by photoactivated L-852,505. As L-852,505 is a transition state analogue, it is expected to bind selectively to active forms of γ -secretase. Thus, the simplest explanation for the failure of L-852,505 to covalently label intact wild-type PS1 is that this form of PS1 is an inactive zymogen.

The binding of L-852,505 to a PS1 variant, PS1 Δ E9, that is associated with early-onset familial Alzheimer's disease¹⁷ was also evaluated. PS1 Δ E9 lacks the purported cytosolic loop encoded by exon 9 which contains the endoproteolytic cleavage site in PS1 (which gives rise to PS1-NTF and PS1-CTF); hence, PS1 Δ E9 is a stable single-chain derivative ($M_r \sim 43K$). Solubilized γ -secretase was prepared from N2a cells that overexpress PS1 Δ E9 (N2a/PS1 Δ E9 solubilized γ -secretase). SDS-PAGE/immunoblot analysis of N2a/PS1 Δ E9 solubilized γ -secretase with antibodies against PS1-NTF and PS1-CTF confirmed the presence of PS1 Δ E9 (Fig. 4b). The absence of PS1 immunoreactivity at the positions expected for PS1-NTF or PS1-CTF reflects the so-called 'replacement' phenomenon in which overexpression of human PS1 Δ E9 results in compromised accumulation of the murine PS1/PS2 heterodimers¹⁸. Photoactivation of L-852,505 in the presence of N2a/PS1 Δ E9 solubilized γ -secretase yields a covalently labelled biotinylated protein that migrates with PS1 Δ E9 and reacts with both the anti-PS1-NTF and anti-PS1-CTF antibodies (Fig. 4b). The appearance of this biotinylated protein is blocked when the photoactivation is performed in the presence of excess L-685,458. These results suggest that PS1 Δ E9 solubilized γ -secretase also binds transition state analogue inhibitors of aspartyl proteases and is thus catalytically competent. Consistent with this view, N2a/PS1 Δ E9 solubilized γ -secretase displays activity in the *in vitro* γ -secretase assay and L-685,458 blocks this activity with a potency similar to that observed towards wild-type solubilized γ -secretase (data not shown). The fact that single-chain PS1 Δ E9 is associated with catalytic activity,

whereas intact single-chain wild-type PS1 is inactive (as it cannot bind the photoaffinity probes), indicates that the putative activating conformational change triggered by endoproteolytic cleavage of intact wild-type PS1 may be mimicked by loss of the cytosolic loop in PS1 Δ E9.

Additional insight into the selectivity of the L-852,505 photoaffinity probe was gleaned from our studies with N2a/PS1 Δ E9 solubilized γ -secretase. Biotinylated species that migrate with P20 or P23 were not observed after treatment of N2a/PS1 Δ E9 solubilized γ -secretase with photoactivated L-852,505 (Fig. 4b). This result indicates that there is little or no covalent labelling of proteins other than PS1-CTF and PS2-CTF when wild-type solubilized γ -secretase is photolabelled with L-852,505—assuming that overexpression of PS1 Δ E9 only suppresses (through the 'replacement' phenomenon) the appearance of the PS1 and PS2 fragments.

Our results provide compelling biochemical evidence that PS1 contains the active site of γ -secretase, assuming that our active-site-directed photoreactive probes covalently label the protein containing the active site of γ -secretase rather than a contiguous protein. This assumption is supported by the finding that two different photoaffinity analogues, L-852,505 and L-852,646, covalently label the same protein—PS1. The fact that L-852,505 and L-852,646 bind to dissimilar PS1 fragments indicates that the active site of γ -secretase may be shared between the two subunits. This prediction agrees with that derived from the PS1 mutagenesis study in which the aspartates in the putative active site in PS1 are situated on dissimilar subunits⁸.

The conclusion that PS1 contains the active site of γ -secretase is consistent with results from our biochemical studies showing that immunodepletion of PS1 from solubilized γ -secretase causes a concomitant reduction in γ -secretase activity, and that solubilized γ -secretase activity migrates with PS1 during gel exclusion chromatography¹². Although PS1 appears to contain the active

site of γ -secretase, it does not exclude involvement of other protein cofactors as determinants of γ -secretase activity. Our suggestion that γ -secretase activity is catalysed by a PS1-containing macromolecular complex is consistent with the existence of such putative cofactors¹². Further studies focused on dissociation and reassembly of the macromolecular γ -secretase complex should elucidate the role in catalysis of the PS1 fragments and other protein cofactors. □

Methods

Chemical compounds

We identified (1S-benzyl-4R-(1-(1S-carbamoyl-2-phenylethylcarbamoyl)-1S-3-methyl-butylcarbamoyl)-2R-hydroxy-5-phenylpentyl) carbamic acid *tert*-butyl ester (L-685,458) as an inhibitor of cellular A β secretion by directed screening of the Merck sample collection⁹. L-685,458 and ((1S-benzyl-4R-(1-(1S-carbamoyl-2-phenylethylcarbamoyl)-1S-3-methyl-butylcarbamoyl)-2S-hydroxy-5-phenylpentyl) carbamic acid *tert*-butyl ester) (L-682,679) will be described elsewhere⁹. tBoc-Phe-Phe-Leu-Phe-amide (L-405,484) was synthesized by Enzyme Systems Products. Synthesis of (4R-(1S-(2-(4-benzoyl-phenyl)-1-(5-(5-(2-oxo-(3aR,6aS)hexahydro-thieno(3,4-d)imidazol-6S-yl)-pentanoylamino)-pentylcarbamoyl)-ethylcarbamoyl)-3-(1S)-methyl-butylcarbamoyl)-1S-benzyl-2R-hydroxy-5-phenyl-pentyl)-carbamoyl carbamic acid *tert*-butyl ester (L-852,505) was performed as follows. A solution of BOC-Leu-BPA-OH (29 mg), EDC (11 mg), HOBT (8 mg), triethylamine (9 μ L) in DMF (1 ml) was treated with 5-(biotinamido)pentylamine TFA (30 mg) and stirred at room temperature for 60 min. The reaction mixture was diluted with ethyl acetate, washed with aqueous citric acid, NaHCO₃ and brine, then dried (MgSO₄). The crude reaction product was purified by chromatography on silica gel (eluant, 8% MeOH-CH₂Cl₂; yield, 30 mg, 63%). This product was treated with TFA (1 ml) at room temperature for 10 min, evaporated under vacuum and purified by chromatography on silica gel (eluant: 10% (1:10 NH₄OH/MeOH)-CH₂Cl₂). A solution of the resulting product (69 mg) in DMF (1 ml) was treated with 2R-benzyl-5S-*tert*-butoxycarbonylamino-4R-(*tert*-butyldimethylsilyloxy)-6-phenyl-hexanoic acid (53 mg) (ref. 19), EDC (22 mg) and HOBT (15 mg) and stirred overnight at room temperature. The reaction mixture was diluted with ethyl acetate, washed with aqueous citric acid, NaHCO₃ and brine, then dried (MgSO₄). The crude reaction product was purified by chromatography on silica gel (eluant, 8% MeOH-CH₂Cl₂; yield, 101 mg, 84%). The product of this reaction was treated with TBAF (1.0 M in THF, 1 ml) and stirred overnight at room temperature. The reaction mixture was diluted with aqueous citric acid, water and ether. The resulting precipitate was collected by filtration, washed thoroughly with water and ether and dried under vacuum (yield 52 mg, 57%). Data for L-852,505: δ (1H, DMSO) 7.90–7.02 (24H, m), 6.45 (1H, d, J = 9.0), 6.38 (1H, s), 6.34 (1H, s), 4.80–4.00 (5H, m), 3.50–2.45 (16H, m), 2.01 (apparent t, J = 7.2, 2H), 1.67–0.75 (22H, m), m^+/z 1089 = (C₆₁H₈₁N₇O₉S)H⁺. We synthesized 4-benzoyl-N-(1S-benzyl-2R-hydroxy-4R-(3-methyl-1S-(1-(5-(5-(2-oxo-(3aR,6aS)hexahydro-thieno(3,4-d)imidazol-6S-yl)-pentanoylamino)-pentylcarbamoyl)-2S-phenyl-ethylcarbamoyl)-butylcarbamoyl)-5-phenyl-pentyl)-benzamide (L-852,646) in a similar manner. Data for L-852,646: δ (1H, DMSO) 8.34 (1H, d, J = 9.0), 7.91–7.69 (11H, m), 7.60–7.57 (2H, m), 7.25–7.11 (16H, m), 6.41 (1H, s), 6.35 (1H, s), 4.94 (1H, d, J = 6), 4.44–4.12 (5H, m), 3.62–3.58 (1H, m), 3.19–2.56 (13H, m), 2.03 (2H, apparent t, J = 7), 1.73–0.9 (17H, m), 0.79 (3H, d, J = 6.5), 0.74 (3H, d, J = 6.5), m^+/z 1093 = (C₆₃H₇₇N₇O₈S)H⁺.

Electrochemiluminescence assay for the A β peptides

The A β peptides were detected by electrochemiluminescence assays^{20,21} using a variety of anti-A β antibodies and an Origen 1.5 Analyzer (Igen Inc.). The 4G8 murine monoclonal antibody (Senetek PLC) binds to an epitope in the A β peptide (within amino acids 18–21) that is immediately distal to the α -secretase cleavage site. The G2-10 murine monoclonal antibody (provided by K. Beyreuther) binds to the C terminus that is exposed after the γ -secretase-mediated cleavage which generates amino acid 40 of the A β peptide²². The FCA3542 rabbit antibody (provided by F. Checler) binds to the C terminus that is exposed after the γ -secretase-mediated cleavage that generates amino acid 42 of the A β peptide²³. The 4G8 monoclonal antibody was biotinylated with Biotin-LC-Sulfo-NHS-Ester (Igen Inc.). The G2-10 and FCA3542 antibodies were ruthenylated with TAG-NHS Ester (Igen Inc.). A β (x-40) was detected with biotinylated 4G8 and ruthenylated G2-10. A β (x-42) was detected with biotinylated 4G8 and ruthenylated FCA3542.

Membrane preparation and detergent solubilization

Frozen HeLa S3 cells were resuspended in buffer A (50 mM MES, pH 6.0, 5 mM MgCl₂, 5 mM CaCl₂, 150 mM KCl) spiked with 'complete' protease inhibitor cocktail tablets (Boehringer Mannheim). The cells were broken by single-pass through a French press (Spectronic Instruments). Cell debris and nuclei were removed by centrifugation at 800g for 10 min. The supernatants were centrifuged at 100,000g for 60 min. The 100,000g pellets were resuspended in buffer A and the centrifugation was repeated. The final membrane pellets were resuspended in buffer A to a protein concentration of $\sim$ 2 mg ml⁻¹. All procedures were performed at 4 °C. The membranes were stored at -70 °C. Detergent solubilization of HeLa cell membranes (protein concentration, 5 mg ml⁻¹ in buffer A) involved treatment with 1% CHAPSO for 60 min at 4 °C and centrifugation at 100,000g for 60 min. The supernatant fraction is defined as solubilized γ -secretase. Similarly, membrane extracts were prepared from N2a neuroblastoma cells stably transfected with PS1 and the APP695-Swedish variant and N2a cells stably transfected with the PS1-FAD Δ exon 9 mutant and the APP695-Swedish variant (provided by S. Sisodia).

In vitro γ -secretase assay

A DNA fragment encoding amino acids 596–695 of APP and the Flag sequence (DYKDDDDK) at the C terminus was generated by PCR amplification with suitably designed oligonucleotides and a plasmid that contained APP695. The Met that serves as the translation start site is residue 596 of APP (the P1 residue with respect to the β -secretase cleavage site). This DNA fragment was inserted into the prokaryotic expression vector pET2-21b (Novagen). The recombinant protein, C100Flag, was overproduced in *Escherichia coli* (strain BL21(DE3)) and purified by Mono-Q column chromatography (Pharmacia Biotechnology). C100Flag (1.7 μ M) was incubated with solubilized γ -secretase (125 μ g ml⁻¹) in the presence of CHAPSO (0.25%) in buffer B (50 mM PIPES, pH 7.0, 5 mM MgCl₂, 5 mM CaCl₂, 150 mM KCl) at 37 °C. The reactions were stopped by adding RIPA (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris HCl, pH 8.0) and boiling for 5 min. The samples were centrifuged and the supernatant solutions were assayed for the A β peptide by electrochemiluminescence.

Covalent binding of photoreactive ligands

Solubilized γ -secretase (2.5 mg ml⁻¹) was mixed with streptavidin-agarose beads (Pierce Chemical Co.) at 4 °C for 60 min, followed by centrifugation to remove endogenous biotinylated species. The unbound fraction was diluted 10-fold in buffer B and incubated with 10–25 nM of the photoaffinity ligands, L-852,505 and L-852,646. In parallel, samples were prepared in the presence of L-685,458 (2 μ M) to determine non-specific binding of the photoreactive ligands. The samples were irradiated on ice for 45 min using a Rayonet RPR-200 photochemical reactor with PRP-3500A lamps (Southern New England Ultraviolet Co.). The photoactivated samples were mixed with 0.2 volume of 5 $\times$ RIPA and incubated with streptavidin-agarose beads overnight at 4 °C. The streptavidin-agarose beads were washed three times with RIPA and the proteins were eluted with SDS sample buffer and boiled for 5 min. The samples were centrifuged and the supernatant solutions were fractionated by SDS-PAGE and electrophoretically transferred to Immobilon-P membranes (Millipore). The blots were probed for (1) biotin with affinity-purified rabbit polyclonal anti-biotin antibody (Pierce Chemical Co.), (2) PS1-CTF with rabbit polyclonal antibody Ab29498 (provided by K. Beyreuther & C. Elle) (3) PS1-NTF with affinity purified rabbit polyclonal Ab14A (provided by S. Gandy) and (4) PS2-CTF with affinity purified rabbit polyclonal antibody PC235 (Oncogene Research Products).

Received 22 February; accepted 28 April 2000.

- Selkoe, D. J. The cell biology of β -amyloid precursor protein and presenilin in Alzheimer's disease. *Trends Cell Biol.* **8**, 447–453 (1998).
- Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754–760 (1995).
- Levy-Lahad, E. *et al.* A familial Alzheimer's disease locus on chromosome 1. *Science* **269**, 970–973 (1995).
- Scheuner, D. *et al.* Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Med.* **2**, 864–870 (1996).
- Duff, K. *et al.* Increased amyloid-beta42(43) in brains of mice expressing mutant presenilin 1. *Nature* **383**, 710–713 (1996).
- Borchelt, D. R. *et al.* Familial Alzheimer's disease-linked presenilin 1 variants elevate Abeta42, 1-40 ratio in vitro and in vivo. *Neuron* **17**, 1005–1013 (1996).
- De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**, 387–390 (1998).
- Wolfe, M. S. *et al.* Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ -secretase activity. *Nature* **398**, 513–517 (1999).
- Shearman, M. S. *et al.* L-685,458, an aspartic protease transition state mimic, is a potent inhibitor of APP γ -secretase activity. *Biochemistry*, in the press.
- Wolfenden, R. Analog approaches to the structure of the transition state in enzyme reactions. *Accs. Chem. Res.* **5**, 10–18 (1972).
- Rich, D. H. *Inhibitors of Aspartic Proteinases*. (eds. Barrett, A. J. & Salvesen, G.) 179–217 (Elsevier Science, 1986).
- Li, Y.-M. *et al.* Presenilin-1 is linked with γ -secretase activity in the detergent solubilized state. *Proc. Natl Acad. Sci. USA* (in the press).
- Wolfe, M. S. *et al.* Peptidomimetic probes and molecular modeling suggest that Alzheimer's γ -secretase is an intramembrane-cleaving aspartyl protease. *Biochemistry* **38**, 4720–4727 (1999).
- Thinakaran, G. *et al.* Endoproteolysis of presenilin-1 and accumulation of processed derivatives in vivo. *Neuron* **17**, 181–190 (1996).
- Seeger, M. *et al.* Evidence for phosphorylation of oligomeric assembly of presenilin 1. *Proc. Natl Acad. Sci. USA* **94**, 5090–5094 (1997).
- Capell, A. *et al.* The proteolytic fragments of the Alzheimer's disease-associated presenilin-1 form heterodimers and occur as a 100–150-kDa molecular mass complex. *J. Biol. Chem.* **273**, 3205–3211 (1998).
- Perez-Tur, J. *et al.* A mutation in Alzheimer's disease destroying a splice acceptor site in the presenilin-1 gene. *Neuroreport* **7**, 297–301 (1995).
- Thinakaran, G. *et al.* Evidence that levels of presenilins (PS1 and PS2) are coordinately regulated by competition for limiting cellular factors. *J. Biol. Chem.* **272**, 28415–28422 (1997).
- Evans, B. E. *et al.* A stereocontrolled synthesis of hydroxyethylene dipeptide isosteres using novel, chiral aminoalkyl epoxides and γ -(aminoalkyl) γ -lactones. *J. Org. Chem.* **50**, 4615–4625 (1985).
- Yang, H., Leland, J. K., Yost, D. & Massey, R. J. Electrochemiluminescence: A new diagnostic and research tool. *BioTechnology* **12**, 193–194 (1994).
- Khorkova, O. E., Patel, K., Heroux, J. & Sahasrabudhe, S. Modulation of amyloid precursor protein processing by compounds with various mechanisms of action: detection by liquid phase electrochemiluminescent system. *J. Neurosci. Methods* **82**, 159–166 (1998).
- Ida, N. *et al.* Analysis of heterogeneous BA4 peptides in human cerebrospinal fluid and blood by a newly developed sensitive western blot analysis. *J. Biol. Chem.* **271**, 22908–22914 (1996).
- Barelli, H. L. *et al.* Characterization of new polyclonal antibodies specific for 40 and 42 amino

acid-long amyloid β -peptides: Their use to examine the cell biology of presenilins and the immunohistochemistry of sporadic Alzheimer's disease and cerebral amyloid angiopathy cases. *Mol. Med.* 3, 695–707 (1997).

Correspondence and requests for materials should be addressed to Y.-M.L. (e-mail: yueming_li@merck.com) or S.J.G. (e-mail: steve_gardell@merck.com).

α -Haemolysin of uropathogenic *E. coli* induces Ca^{2+} oscillations in renal epithelial cells

Per Uhlén[†], Åsa Laestadius[†], Timo Jahnukainen^{*}, Tomas Söderblom[‡], Fredrik Bäckhed[‡], Gianni Celsi^{*}, Hjalmar Brismar^{*}, Staffan Normark[‡], Anita Aperia^{*} & Agneta Richter-Dahlfors[‡]

^{*} Department of Women and Child Health, Karolinska Institutet, Astrid Lindgren Children's Hospital, Q2:09, S-171 76 Stockholm, Sweden

[‡] Microbiology and Tumour Biology Centre, Karolinska Institutet, S-171 77 Stockholm, Sweden

[†] These authors contributed equally to this work

Pyelonephritis is one of the most common febrile diseases in children. If not treated appropriately, it causes irreversible renal damage and accounts for a large proportion of end stage renal failures¹. Renal scarring can occur in the absence of inflammatory cells, indicating that bacteria may have a direct signalling effect on renal cells². Intracellular calcium ($[\text{Ca}^{2+}]_i$) oscillations can protect cells from the cytotoxic effects of prolonged increases in intracellular calcium^{3,4}. However, no pathophysiologically relevant protein that induces such oscillations has been identified. Here we show that infection by uropathogenic *Escherichia coli* induces a constant, low-frequency oscillatory $[\text{Ca}^{2+}]_i$ response in target primary rat renal epithelial cells induced by the secreted RTX (repeats-in-toxin) toxin α -haemolysin. The response depends on calcium influx through L-type calcium channels as well as from internal stores gated by inositol triphosphate. Internal calcium oscillations induced by α -haemolysin in a renal epithelial cell line stimulated production of cytokines interleukin (IL)-6 and IL-8. Our findings indicate a novel role for α -haemolysin in pyelonephritis: as an inducer of an oscillating second messenger response in target cells, which fine-tunes gene expression during the inflammatory response.

To identify the renal cell type targeted by pyelonephritogenic *E. coli* strain ARD6 *in vivo*, we used a model in which pyelonephritis is induced in young rats by intraurethral infusion of bacteria. We localized bacteria by confocal microscopy of immunostained sections of dissected kidneys⁵. Early in infection, bacteria were located at the proximal tubules, where they formed microcolonies on the luminal surface of proximal tubule cells (Fig. 1a). Addition of a derivative of ARD6 expressing green fluorescent protein (GFP) to a primary preparation of rat proximal tubule (RPT) cells confirmed the extracellular location of bacteria at the initial phase of infection (Fig. 1b).

We used primary cultures of RPT cells to study the effect of *E. coli* on signalling mechanisms. We recorded $[\text{Ca}^{2+}]_i$ in single cells loaded with Fura-2/AM. The fluorescence pattern revealed that addition of uropathogenic *E. coli* strain ARD6/pGFP induced oscillations in $[\text{Ca}^{2+}]_i$ (Fig. 1c), whereas control medium did not. The oscillatory response began about 30 min after the cells were exposed to bacteria and was sustained for at least 60 min. In separate experiments, oscillations were observed for as long as 2.5 h (data not shown).

Immunohistochemistry studies and gentamicin assays showed no

evidence of bacterial attachment and intracellular proliferation at this early phase of infection (30 min after infection; data not shown). Therefore, we tested the sterile supernatant from an overnight culture of strain ARD6. Addition of diluted supernatant induced an oscillatory calcium response in RPT cells within 2–5 min (Fig. 1d). When the diluted supernatant was replaced by control medium, immediate inhibition of the calcium response occurred. This indicates that the continuous presence of the inducer

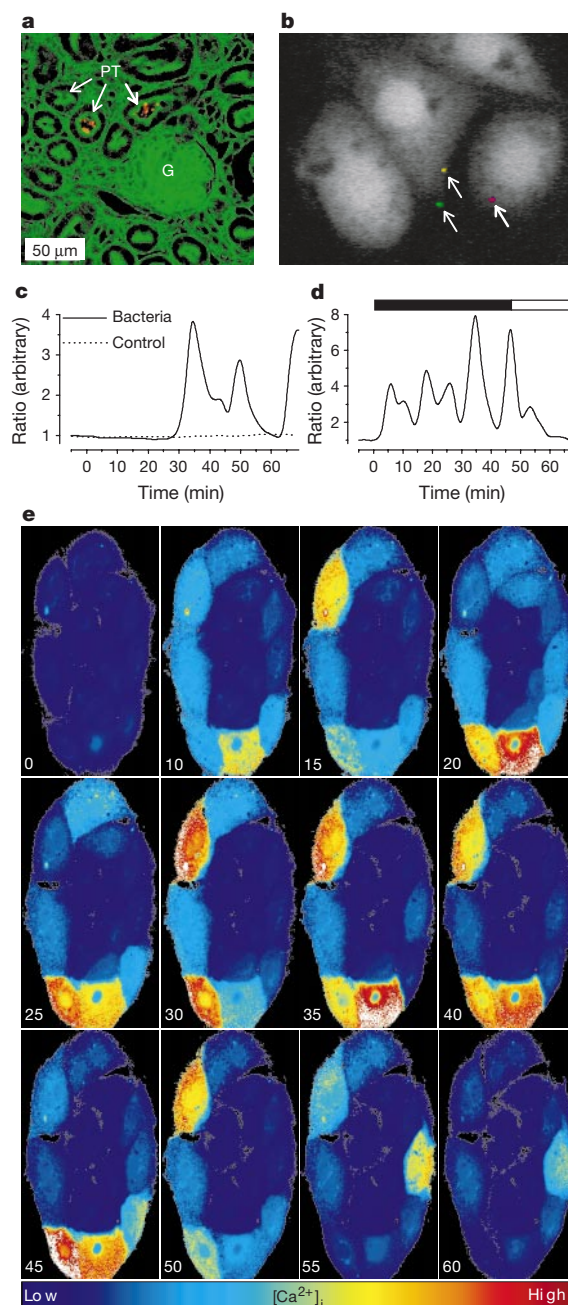


Figure 1 Bacteria target proximal tubules and induce oscillations of $[\text{Ca}^{2+}]_i$ in primary RPT cells. **a**, Localization of bacteria (red) in renal parenchyma (outlined by FITC–phalloidin, green) visualized by confocal laser scanning microscopy. PT, proximal tubule; G, glomerulus. **b**, Movement of one bacterium at 20-s intervals (green, red and yellow) five min after bacterial addition. **c**, A representative single cell tracing of $[\text{Ca}^{2+}]_i$ following stimulation with ARD6 or control medium. **d**, A single cell tracing of $[\text{Ca}^{2+}]_i$ in the presence (black bar) and absence (white bar) of supernatant from strain ARD6. **e**, $[\text{Ca}^{2+}]_i$ in individual cells at different times (min) after stimulation by supernatant from ARD6.

is required for oscillations to occur. Images of the oscillatory response are shown in Fig. 1e. The oscillatory response is observed in virtually all cells located in the periphery of the cell cluster.

A spectral analysis⁶ indicates that a secreted factor from *E. coli* strain ARD6 induces calcium oscillations with a constant frequency (Fig. 2). Only cells that showed a typical oscillatory response were included in the analysis. The dominant peak occurred at 1.4 ± 0.1 mHz (11 experiments, 80 cells) which corresponds to a periodicity of 12.0 ± 0.7 min. The $[Ca^{2+}]_i$ increased from a basal level of 0.1 ± 0.005 μ M to 0.5 – 1.5 μ M. Calcium oscillations with a periodicity in the minute range have previously been demonstrated in both excitable and non-excitable cells^{7,8}.

The $[Ca^{2+}]_i$ response was abolished by blocking voltage-operated L-type calcium channels with nifedipine⁹ (Fig. 3a). The calcium response was also completely blocked by 2-APB (2-aminoethoxydiphenyl borate) (Fig. 3b), a cell-permeable inositol triphosphate ($InsP_3$) receptor antagonist¹⁰. Ionomycin, a calcium ionophore, was used to investigate whether an oscillatory $[Ca^{2+}]_i$ response could be reproduced by insertion of calcium-permeable pores within the plasma membrane. This procedure resulted in a persistent rise in $[Ca^{2+}]_i$, but no oscillations (data not shown). Activation of voltage-operated L-type calcium channels with Bay K 8644 or by depolarization through an increase of the $[K^+]$ in the medium to 50 mM, also resulted in a persistent rise in $[Ca^{2+}]_i$, but not oscillations (data not shown). When cells were pre-incubated with the phospholipase C (PLC) inhibitor U73122, and then exposed to the ARD6 supernatant, perturbations in $[Ca^{2+}]_i$ were still observed (Fig. 3c), but the character of the response differed from that observed in cells exposed to the supernatant alone (Fig. 3d). Fewer of the cells pre-incubated with U73122 responded with typical oscillations. In addition, a number of U73122-treated cells exhibited a slow continuous increase in $[Ca^{2+}]_i$. This response was not observed in cells treated with supernatant alone. The data indicate that activation of PLC is at least partly involved in triggering the effects of the bacterial factor on $InsP_3$ receptors. These results indicate that the factor secreted by bacteria does not permeate the plasma membrane, but rather interacts with it and activates both voltage-operated L-type

calcium channels and $InsP_3$ receptors, which then give rise to slow calcium oscillations.

To test whether calcium oscillation is a general response to clinical isolates of uropathogenic *E. coli*, we tested other strains. Supernatant from the pathogenic strain DS17 induced an oscillatory calcium response (Fig. 3e), whereas supernatant from a non-pathogenic *E. coli* K-12 strain, W3110, was incapable of generating a response (Fig. 3f). Pyelonephritogenic *E. coli* isolates frequently carry one or more *pap* pathogenicity islands encoding different virulence determinants¹¹. One such virulence factor is the exotoxin α -haemolysin¹², which is expressed by strains ARD6 and DS17, but not W3110. To test whether α -haemolysin induces the oscillatory calcium response, an α -haemolysin-expressing plasmid (pANN 202–812) was introduced into strain W3110. The supernatant of this culture tested positive for haemolytic activity on sheep erythrocytes (data not shown) and induced an oscillatory calcium response in RPT cells (Fig. 3g). Addition of 80 haemolytic units (HU) of α -haemolysin purified from ARD6 culture supernatant¹³ (equivalent to 10μ l ml⁻¹ of ARD6 supernatant) resulted in typical calcium oscillations (Fig. 3h). No lysis was observed in cells incubated for as long as 24 h. Addition of 800 HU of purified α -haemolysin resulted in a sustained increase in $[Ca^{2+}]_i$ followed by cell lysis within 30 min (Fig. 3i). These data show that secreted α -haemolysin induces the oscillatory calcium response. Binding of

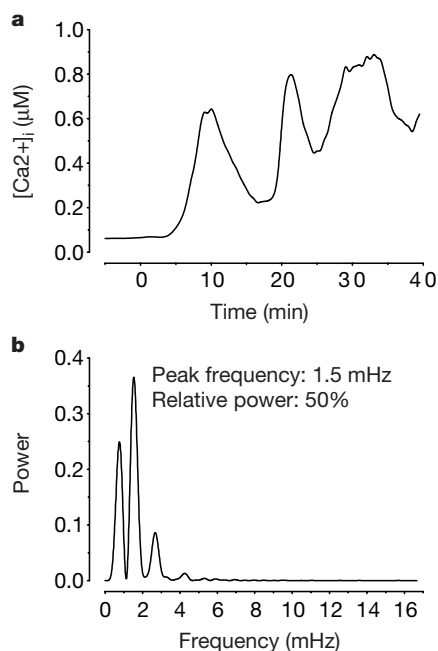


Figure 2 Spectral analysis of supernatant-induced $[Ca^{2+}]_i$ oscillations. **a**, A single cell recording of $[Ca^{2+}]_i$ oscillation and **b**, the corresponding power spectrum analysis. Oscillations occur with a frequency of 1.5 mHz, corresponding to a period of 12.0 ± 0.7 min.

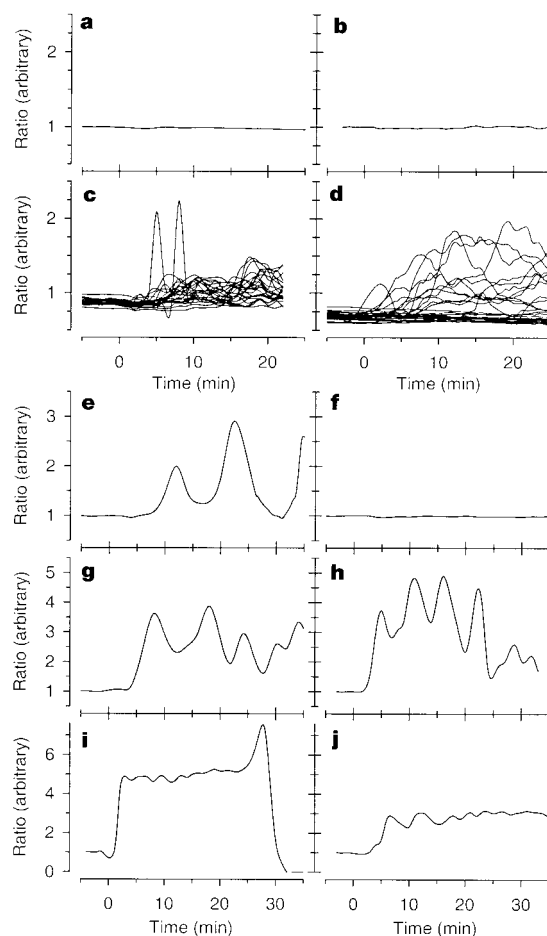


Figure 3 Characterization of the oscillatory $[Ca^{2+}]_i$ response in RPT cells. $[Ca^{2+}]_i$ response in cells pre-incubated with **a**, nifedipine, **b**, 2-APB or **c**, U73122 before addition of supernatant from ARD6. $[Ca^{2+}]_i$ response in cells stimulated by supernatants from **d**, ARD6, **e**, DS17; **f**, W3110; **g**, W3110/pANN202–812; or **h**, 80 HU and **i**, 800 HU of α -haemolysin. **j**, ARD6-stimulated Ca^{2+} -response in the presence of polymyxin B. Each experiment was repeated at least three times using individual cell preparations. Each panel shows one characteristic single cell experiment, except for **c** and **d**, which show the response from ~25 cells.

α -haemolysin to target membrane is vital for haemolytic activity *in vitro*¹⁴. We tested the effects of a strain that expresses and secretes a non-acylated form of α -haemolysin that is defective in its interaction with the plasma membrane¹⁵. Supernatant from this strain had no effect on $[Ca^{2+}]_i$ (data not shown), indicating that an appropriate interaction between α -haemolysin and the plasma membrane is required for calcium oscillations to occur.

Bacterial lipopolysaccharide (LPS) is frequently shed into culture supernatants, and it can also bind in trace amounts to purified proteins. To rule out LPS as a contributory factor for induction of the calcium response, culture supernatant from strain ARD6 or purified α -haemolysin was pre-incubated with polymyxin B before being added to RPT cells. This treatment, which specifically neutralizes LPS, had no effect on the oscillatory response (Fig. 3j). In contrast, heat treatment of ARD6 supernatant and purified α -haemolysin inhibited the effect (data not shown). Thus, calcium oscillations are specifically induced by α -haemolysin, not LPS.

Women who suffer from acute pyelonephritis frequently show elevated urine and serum concentrations of the pro-inflammatory cytokines IL-6 and IL-8 (ref. 16). To investigate the role of α -haemolysin-induced calcium oscillations we measured the release of IL-6 and IL-8 from A498 cells, a human renal epithelial cell line that responds to supernatant from ARD6 by $[Ca^{2+}]_i$ oscillations (Fig. 4a). In contrast to cells exposed to control medium or to supernatant from W3110 (Hly⁻), addition of supernatant from strain ARD6 (Hly⁺) or purified α -haemolysin (80 HU) stimulated IL-6 and IL-8 release (Fig. 4b, c). Pre-treatment of cells with nifedipine abrogated these effects. As expected, cells exposed to IL-1 β exhibited a huge increase in IL-6 and IL-8 release. The increase in IL-8 release was the same in the presence of nifedipine, and the increase in IL-6 release was moderately attenuated. Interestingly, the expression of IL-8 in Jurkat cells has been shown to be mediated by artificially induced Ca^{2+} oscillations^{8,17}. Taken together, our data show that increased

pro-inflammatory cytokine release is a specific effect of α -haemolysin-induced $[Ca^{2+}]_i$ oscillations.

Although elevations of $[Ca^{2+}]_i$ are required for calcium to act as a signal, prolonged increases in $[Ca^{2+}]_i$ can be lethal. To avoid the cytotoxic effects of high $[Ca^{2+}]_i$, cells use either low-amplitude Ca^{2+} signals or, more commonly, calcium oscillations^{1,2}. The role of α -haemolysin and other RTX toxins in inflammation was for a long time believed to depend on their cytolytic activities. This study shows that bacteria are localized extracellularly in the renal tubular lumen during the early phase of disease where they are constantly exposed to the flushing effects of urine. It is therefore unlikely that levels of α -haemolysin capable of causing sustained increases in $[Ca^{2+}]_i$ and cell lysis are achieved *in vivo*. Instead, *E. coli* α -haemolysin contributes to inflammation by inducing low-frequency oscillatory Ca^{2+} signals in renal epithelial cells, resulting in the release of IL-6 and IL-8. Accordingly, we propose that sub-lytic concentrations of this toxin can induce a second messenger response in host cells that results in the induction of inflammation and probably other, yet to be identified cellular events. □

Methods

Preparation of bacteria, supernatant and α -haemolysin

The human and rat pyelonephritogenic *E. coli* strain ARD6 (serotype O6:K13:H1, World Health Organization designation Su 4344/41)¹⁸ was cultured overnight in Luria-Bertani (LB) medium at 37 °C. The same culture conditions were used for all bacteria unless otherwise stated. Other strains used were ARD6/pGFPmut2 (ref. 19) (cultured in the presence of ampicillin (100 μ g ml⁻¹) (Sigma)), DS17 (ref. 20) and W3110 (ref. 21). Washed bacteria were added to the primary RPT cell cultures at 10⁵ CFU per ml. Strain W3110 harbouring plasmids pANN202-812 (ref. 22) (Hly⁺) or pANN202-812B (ref. 23) (Hly⁻) were grown in the presence of ampicillin (100 μ g ml⁻¹) (Sigma), and chloramphenicol (30 μ g ml⁻¹) (Sigma). When bacterial growth supernatant was added, 10 μ l ml⁻¹ of filtered supernatant (Supor Acrodisc 0.2 μ m, Gelman Sciences) was used. α -Haemolysin was purified and assayed for haemolytic activity as described¹³. 1,000 haemolytic units (HU) were defined as the dilution giving 60% haemolysis of a 2.5% sheep erythrocyte suspension.

In vivo model and microscopy of immunostained sections

The bladders of six anaesthetized, female, 20-day-old Sprague-Dawley rats (B&K Universal AB) were catheterized and 0.3 ml (1 $\times$ 10⁶ CFU per ml in PBS) of *E. coli* strain ARD6 was slowly infused. Four days after infection, rats were killed and kidneys prepared for immunostaining and confocal microscopy⁵. The antibodies used were OK K13 antiserum (Statens Serum Institut) and donkey anti-rabbit Texas Red (Jackson ImmunoResearch Labs). Actin was detected with FITC-conjugated phalloidin (Molecular Probes). Samples were observed with a Multiprobe 2001 confocal laser scanning microscope (Molecular Dynamics).

Preparation of primary proximal tubule cells

We used kidneys from uninfected, 20-day-old female Sprague-Dawley rats to prepare proximal tubule cells²⁴. Cells were cultured in supplemented DMEM (20 mM Hepes, 24 mM NaHCO₃, 10 μ g ml⁻¹ penicillin, 10 μ g ml⁻¹ streptomycin and 10% fetal bovine serum) on glass coverslips for 48–72 h in 5% CO₂ at 37 °C. Cells were starved for serum (1% FBS) and cultured in the absence of antibiotics for 24 h before the experiment.

Ratiometric imaging

Cells were incubated with 30 μ M Fura-2 acetylmethyl ester (Fura-2/AM) (Molecular Probes) in P medium (100 mM NaCl, 4 mM KCl, 20 mM Hepes, 25 mM NaHCO₃, 1 mM CaCl₂, 1.2 mM MgCl₂, 1 mM NaH₂PO₄·H₂O, and 10 mM D-glucose) for 75 min before the $[Ca^{2+}]_i$ measurements. Ratiometric imaging was performed using a heated chamber (FCS2, Biopetech) mounted on a Zeiss Axiovert 135 microscope using a 40 $\times$ /1.4 epifluorescence oil-immersion objective. Emission fluorescence was collected via a GenISys image intensifier system connected to a CCD-camera (MTI CCD72; Dage-MTI) and acquisition software from Invision Corporation. Cells were excited every 30 s, which corresponds to a Nyquist frequency of 16.7 mHz. *In vivo* calibrations were performed as described²⁵ with 10⁻⁵ M Ionomycin (Sigma) and K_d set to 224 nM (Molecular Probes). When indicated, cells were preincubated for 15 min in 10–100 μ M Nifedipine (Sigma) or 50 μ M 2-APB (gift from K. Mikoshiba). Polymyxin B (Sigma) was used at 100 U per ml, Bay K 8644 (Sigma) was used at 10 μ M, and 2 μ M of the PLC inhibitor U73122 (Sigma) was added 2 min before the start of the experiment.

Power spectrum analysis

A power spectrum of a signal is the squared magnitude of its Fourier transform, and describes the contribution to a signal by each of its sine wave components⁶. We used MATLAB to filter and centre the oscillating section of a single cell measurement, and the trend was corrected by computing the Gauss least-square approximation and subtracting the trend from the original data. A fast Fourier transform was used to calculate the discrete

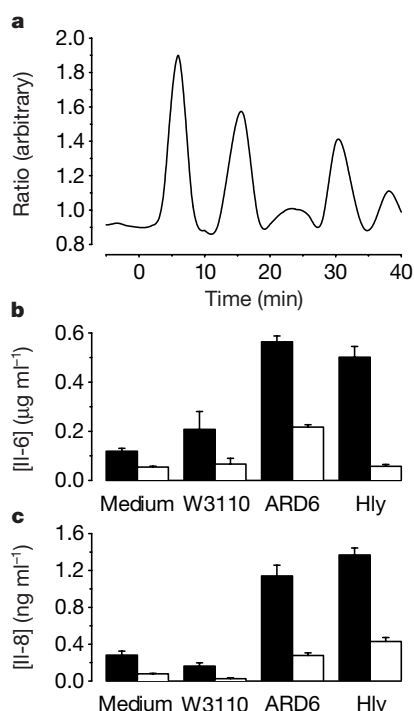


Figure 4 Release of the pro-inflammatory cytokines IL-6 and IL-8 is specifically triggered by α -haemolysin-induced $[Ca^{2+}]_i$ oscillations. **a**, Supernatant from strain ARD6 induces a similar $[Ca^{2+}]_i$ response in A498 cells to that in primary RPT cells. α -haemolysin-induced $[Ca^{2+}]_i$ oscillations stimulate the production of **b**, IL-6 and **c**, IL-8 in A498 cells. Nifedipine (open bars) abrogates these effects. Medium is LB, W3110 and ARD6 indicate the effects of supernatants from these two strains, Hly is purified α -haemolysin, error bars are s.e.m.

Fourier transform. This produces a spectrum where the peaks correspond to the different frequencies present in the original data. The dominant peak was determined by comparing the relative power of the peaks in the spectrum. The relative power was calculated by determining the area between the two extremes closest to the peak, divided by the total area of the power spectrum. Values are presented as standard errors of mean.

Cytokine measurements

The human renal epithelial cell line A498 (ref. 26) (ATCC HTB44) was cultivated in 24-well plates until confluency in RPMI 1640 supplemented with 10% fetal calf serum, 25 mM HEPES and 2 mM L-glutamine (Life Technologies) at 37 °C, 5% CO₂. Cells were washed once, followed by the addition of medium containing supernatant from strain ARD6 (10 µl ml⁻¹). To block voltage-dependent calcium channels, nifedipine (100 µM) was added 15 min before the addition of bacterial supernatant. The tissue culture supernatants were collected after 6 h and analysed for the presence of cytokines IL-6 and IL-8 by ELISA (Diacclone) according to the manufacturer's instructions.

Received 1 February; accepted 14 April 2000.

1. Wing, A. J. in *Oxford Textbook of Clinical Nephrology* Vol. 2 (eds Cameron, S., Davison, A. M., Grünfeld, J.-P., Kerr, D. & Ritz, E.) 1227–1234 (Oxford Univ. Press, New York, 1992).
2. Serlachius, E. et al. Pyelonephritis provokes growth retardation and apoptosis in infant rat renal cortex. *Kidney Int.* **51**, 1855–1862 (1997).
3. Berridge, M. J. Bootman, M. D. & Lipp, P. Calcium—a life and death signal. *Nature* **395**, 645–648 (1998).
4. Berridge, M. J. Elementary and global aspects of calcium signalling. *J. Physiol.* **499**, 291–306 (1997).
5. Richter-Dahlfors, A., Buchan, A. M. J. & Finlay, B. B. Murine salmonellosis studied by confocal microscopy: *Salmonella typhimurium* resides intracellularly inside macrophages and exerts a cytotoxic effect on phagocytes in vivo. *J. Exp. Med.* **186**, 569–580 (1997).
6. Jenkins, G. M. & Watts, D. G. in *Spectral Analysis and its Applications* 1–525 (Holden-Day, San Francisco, 1969).
7. Meldolesi, J. Oscillation, activation, expression. *Nature* **392**, 863–866 (1998).
8. Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392**, 933–936 (1998).
9. Miller, R. J. Multiple calcium channels and neuronal function. *Science* **235**, 46–52 (1987).
10. Maruyama, T., Kanaji, T., Nakade, S., Kanno, T. & Mikoshiba, K. 2APB, 2-aminoethoxydiphenyl borate, a membrane-penetrable modulator of Ins(1,4,5)P₃-induced Ca²⁺ release. *J. Biochem.* **122**, 498–505 (1997).
11. Blum, G. et al. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immunol.* **62**, 606–614 (1994).
12. Hacker, J., Hughes, C., Hof, H. & Goebel, W. Cloned hemolysin genes from *Escherichia coli* that cause urinary tract infections determine different levels of toxicity in mice. *Infect. Immunol.* **42**, 57–63 (1983).
13. Bhakdi, S., Mackman, N., Nicaud, J.-M. & Holland, B. *Escherichia coli* hemolysin may damage target cell membranes by generating transmembrane pores. *Infect. Immunol.* **52**, 63–69 (1986).
14. Ludwig, A., Vogel, M. & Goebel, W. Mutations affecting activity and transport of haemolysin in *Escherichia coli*. *Mol. Gen. Genet.* **206**, 238–245 (1987).
15. Stanley, P., Koronakis, V. & Hughes, C. Aclation of *Escherichia coli* hemolysin: a unique protein lipidation mechanism underlying toxin function. *Microbiol. Mol. Biol. Rev.* **62**, 309–333 (1998).
16. Jacobson, S. H., Hylander, B., Wretling, B. & Brauner, A. Interleukin-6 and interleukin-8 in serum and urine in patients with acute pyelonephritis in relation to bacterial-virulence-associated traits and renal function. *Nephron* **67**, 172–179 (1994).
17. Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca²⁺ response amplitude and duration. *Nature* **386**, 855–858 (1997).
18. Mattsby-Baltzer, I. et al. Experimental *Escherichia coli* ascending pyelonephritis in rats: changes in bacterial properties and the immune responses to surface antigens. *Infect. Immunol.* **35**, 639–646 (1982).
19. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
20. Tullus, K., Horlin, K., Svenson, S. B. & Kallenius, G. Epidemic outbreaks of acute pyelonephritis caused by nosocomial spread of P fimbriated *Escherichia coli* in children. *J. Infect. Dis.* **150**, 728–736 (1984).
21. Bachmann, B. J. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol. Rev.* **36**, 525–557 (1972).
22. Vogel, M., Hess, J., Then, I., Juarez, A. & Goebel, W. Characterization of a sequence (*hlyR*) which enhances synthesis and secretion of hemolysin in *Escherichia coli*. *Mol. Gen. Genet.* **212**, 76–84 (1988).
23. Ludwig, A., Vogel, M. & Goebel, W. Mutations affecting activity and transport of haemolysin in *Escherichia coli*. *Mol. Gen. Genet.* **206**, 238–245 (1987).
24. Larsson, S., Aperia, A. & Lechene, C. Studies on final differentiation of rat renal proximal tubular cells in culture. *Am. J. Physiol.* **251**, C455–464 (1986).
25. Bolsover, S. R., Angus Silver, R. & Whitaker, M. in *Electronic Light Microscopy* (ed. Shotton, D.) 181–210 (Wiley-Liss, New York, 1993).
26. Giard, D. J. et al. *In vitro* cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. *J. Natl Cancer Inst.* **51**, 1417–1423 (1973).

Acknowledgements

We thank E. Zettergren and A.-C. Eklöf for assistance. This work was supported by grants from the programmes 'Infection and Vaccinology' and 'Glycoconjugates in Biological Systems' sponsored by the Swedish Foundation for Strategic Research (to A.R.-D. and F.B.), the Philipsson Foundation (to A.A.), the Medical Research Council (to A.A. and S.N.), the Swedish Cancer Foundation and Bristol-Myer Squibbs (to S.N.), and the Wenner-Gren Foundation (to A.R.-D.).

Correspondence and requests for materials should be addressed to A.R.-D. (e-mail: agneta.richter.dahlfors@mtc.ki.se).

Frequent chromosomal translocations induced by DNA double-strand breaks

Christine Richardson & Maria Jasin

Cell Biology Program, Memorial Sloan-Kettering Cancer Centre and Cornell University Graduate School of Medical Sciences, 1275 York Avenue, New York, New York 10021, USA

The faithful repair of DNA damage such as chromosomal double-strand breaks (DSBs) is crucial for genomic integrity. Aberrant repair of these lesions can result in chromosomal rearrangements, including translocations, which are associated with numerous tumours^{1,2}. Models predict that some translocations arise from DSB-induced recombination in differentiating lymphoid cell types^{3–5} or from aberrant repair of DNA damage induced by irradiation or other agents^{6–8}; however, a genetic system to study the aetiology of these events has been lacking. Here we use a mouse embryonic stem cell system to examine the role of DNA damage on the formation of translocations. We find that two DSBs, each on different chromosomes, are sufficient to promote frequent reciprocal translocations. The results are in striking contrast with interchromosomal repair of a single DSB in an analogous system in which translocations are not recovered. Thus, while interchromosomal DNA repair does not result in genome instability *per se*, the presence of two DSBs in a single cell can alter the spectrum of repair products that are recovered.

In mammalian cells, repair of DSBs occurs by homology-dependent and homology-independent (non-homologous) recombination mechanisms. As a result, DSBs are potent inducers of recombination, increasing both homologous and non-homologous events by three orders of magnitude or more^{9–11}. One pathway of homologous repair, gene conversion—in particular gene conversion not associated with exchange of flanking markers—is a conservative pathway, as sequence information lost at the break

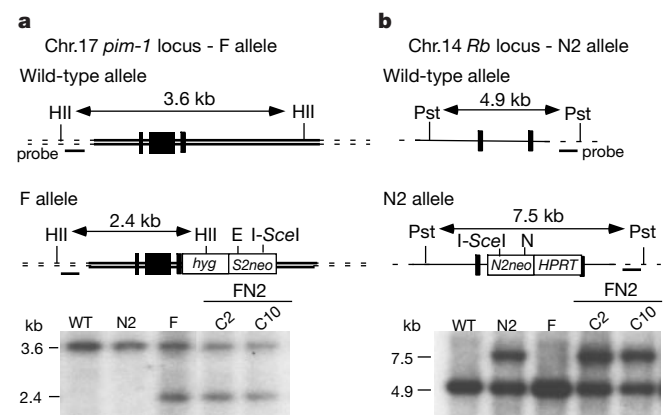


Figure 1 Translocation substrates. The F allele (a) was constructed by targeting *S2neo* and the hygromycin resistance gene (*hyg*) to the *pim-1* locus on chromosome (Chr.) 17 and the N2 allele (b) by targeting *N2neo* and the hypoxanthine phosphoribosyltransferase gene (*HPRT*) to the *Rb* locus on chromosome 14. Wild-type (top) and targeted (bottom) alleles are shown for each locus, with confirmatory Southern analyses presented using *HincII* and *PstI* for the F and the N2 alleles, respectively. *I-SceI* sites disrupted the *S2neo* and *N2neo* genes at *NcoI* (F allele) and *EagI* (N2 allele) sites, respectively, which are 0.53 kb apart. The N allele in the FN cell line is identical to the N2 allele except that it contains a disrupting *PacI* site insertion, rather than an *I-SceI* site²⁹. Black bars, *pim-1* (a) or *Rb* (b) exons. HincII; PstI; N, *NcoI*; E, *EagI*.

site is restored during repair. Sister chromatids are preferred templates for homologous repair by gene conversion^{12,13}. An alternative homologous pathway, single-strand annealing (SSA)¹⁴, occurs predominantly at a DSB between two close direct repeats. Because DNA ends at a DSB are processed to 3' single-stranded tails, complementary single strands formed at direct repeats can anneal, with extruded, non-complementary strands cleaved away. As a result, sequence information between the repeats is lost in the final recombination product¹¹, making SSA a non-conservative type of DSB repair. Non-homologous end-joining (NHEJ) has minimal or no requirement for homology at the site of joining. Like SSA, NHEJ is a non-conservative repair pathway, as nucleotides adjacent to the DSB can be lost during the joining of the broken chromosome ends¹⁵.

DSB repair involving two different chromosomes has the potential to promote genome rearrangements, including translocations, although the number of lesions needed to promote these events is controversial³⁻⁸. Also unknown is the role of various repair pathways in limiting these events in mammalian cells. Sequence repeats are frequently found at the site of genome rearrangements¹⁶. However, gene conversion between repeats on two different chromosomes does not lead to rearrangements at a detectable frequency, presumably because it is not usually associated with exchange of flanking markers¹⁷. The outcome of SSA events has not been examined; however, the annealing of complementary single strands from two different chromosomes will necessarily give rise to translocations. In yeast, SSA events between two chromosomes lead to translocations at high frequency, with as many translocations as intrachromosomal SSA events¹⁸. The contribution of NHEJ to chromosomal rearrangements is unclear, as mutations in some yeast genes involved in this pathway do not lead to a substantial accumulation or suppression of rearrangements¹⁹. Nevertheless, from sequence analysis, the translocation breakpoints from leukemic cell lines often appear to be classic NHEJ junctions²⁰, making it imperative to examine the aetiology of these events in mammalian cells.

To examine directly the potential of multiple DSBs to result in genome rearrangements in mammalian cells, we developed an embryonic stem (ES) cell system using the rare-cutting I-SceI endonuclease^{21,22}. We created the cell line FN2 by marking loci on chromosomes 14 and 17 with two defective neomycin phosphotransferase genes (*neo*), each containing the cleavage site for I-SceI (Fig. 1). Homologous recombination between the two *neo* genes, either with or without exchange of flanking markers, can result in formation of a *neo*⁺ gene which confers G418 resistance. As controls, we used the FN cell line, which contains the defective *neo* genes but only the gene on chromosome 17 contains an I-SceI cleavage site¹⁷, and the F and N2 cell lines, which each contain a single defective *neo* gene with an I-SceI cleavage site (chromosomes 17 and 14, respectively; Fig. 1).

Table 1 Summary of DSB-induced recombination

Cell line	Recombination frequency ($\times 10^{-6}$)		
	No DNA	I-SceI + pneo-WT	I-SceI
F	<0.09	210 $\pm$ 110	<0.09
N2	<0.09	280 $\pm$ 56	<0.09
FN (4-15)	<0.09	330 $\pm$ 140	5 $\pm$ 3
FN2 (C2)	<0.09	730 $\pm$ 120	280 $\pm$ 80
FN2 (C10)	<0.09	1,200 $\pm$ 280	520 $\pm$ 360

Targeted cell lines were electroporated with the I-SceI expression vector, pCBASce¹⁷ or, as controls, with no DNA, the internal *neo* fragment pneo-WT¹⁷ (not shown) or both pCBASce and pneo-WT. Two independently targeted FN2 cell lines, C2 and C10, were electroporated in these experiments. Cells were plated at low density immediately following electroporation and allowed to recover overnight, during which time no appreciable cell division occurred. After 20 h, one plate from each sample was used to determine the number of viable cells after transfection. No significant difference in viability was observed between cell lines or experiments. Cells were then selected in G418 for 12 days and colonies scored. Recombination frequency was calculated as total number G418 resistant clones recovered divided by the total number of viable cells per sample. For all sample frequencies and standard deviations, $n = 4$. Samples that did not give rise to any clones are indicated as a frequency of $<0.09 \times 10^{-6}$.

We did not detect any spontaneously occurring *neo*⁺ recombinants in these four cell lines (frequency $< 10^{-7}$; Table 1). Nor did we detect recombinants following electroporation of a *neo* donor plasmid, pneo-WT, which can correct the mutations in the *neo* genes by recombination (data not shown). To induce DSBs, we electroporated the cell lines with an I-SceI endonuclease expression vector. We observed no significant difference in viability between the different cell lines after electroporation. After electroporation, we did not obtain any *neo*⁺ colonies from the F and N2 cell lines, but with both the FN and FN2 cell lines we observed a high frequency of *neo*⁺ colonies. Thus, the presence of both *neo* genes is required for the recovery of detectable levels of recombinants upon DSB induction. The FN cell line, capable of only a single DSB, gave *neo*⁺ recombinants at an average frequency of 5×10^{-6} per viable cell, at least 50-fold more than spontaneous events (Table 1). The FN2 cell lines, capable of two DSBs, gave recombinants at an average frequency of 4×10^{-4} per viable cell (Table 1), a more than 4,000-fold stimulation over the spontaneous level and an 80-fold stimulation of interchromosomal recombination as compared with the FN cell line. As only a twofold increase would be expected in cells

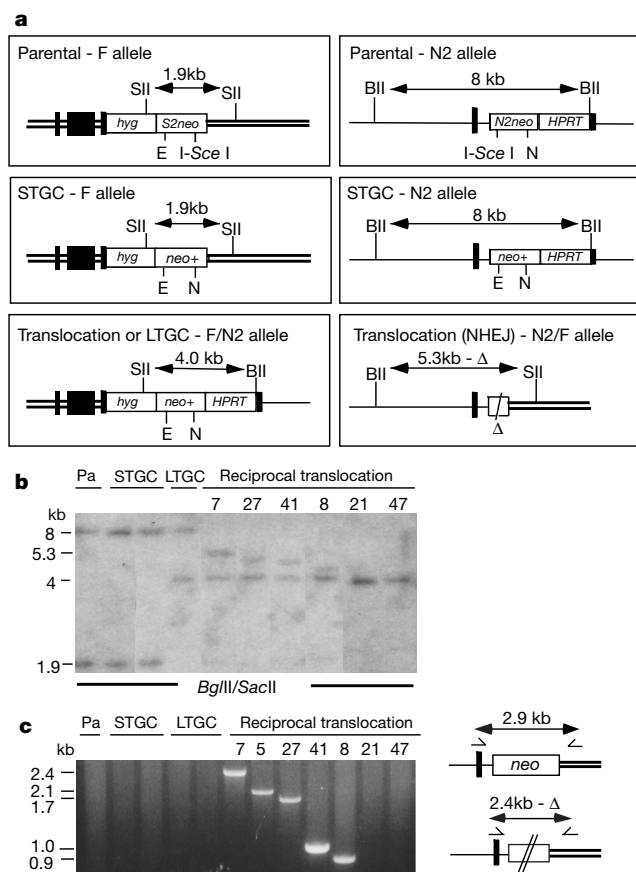


Figure 2 Analysis of *neo*⁺ clones. **a**, Restriction fragment sizes of parental and recombinant alleles. In a STGC event, the I-SceI site is converted to an *Nco*I (F allele) or *Eag*I (N2 allele) site. In a LTGC event, the conversion extends either downstream (F allele) or upstream (N2 allele) of the I-SceI site, forming a hybrid F/N2 allele. As the converted sequence rejoins the original chromosome, the entire chromosome is either F/N2/F or N2/F/N2 (not shown). In a reciprocal translocation, both F/N2 and N2/F hybrid alleles and chromosomes are formed. BII, *Bgl*/II; SII, *Sac*/II. **b**, Southern blot analysis of *neo*⁺ clones. Genomic DNA was digested with *Bgl*/II/*Sac*/II and analysed using a 5' *neo* fragment probe. A 4.0-kb F/N2 hybrid band is present in all translocation clones (numbered across the top). The N2/F hybrid allele ranges in size from 5.3 kb (clone 7) to not detected (clones 21, 47). Pa, parental cell line. **c**, PCR analysis of *neo*⁺ clones. Hybrid N2/F alleles were amplified with primers from exon 20 of *Rb* and intron 4 of *pim*-1.

capable of two DSBs if their effects were independent, this finding suggests a synergistic effect of the second DSB on interchromosomal recombination.

As an additional control, we electroporated cells with both the donor pneo-WT plasmid and the I-SceI expression vector to promote the restoration of *neo* gene function at the F and N2 alleles by DSB-induced homologous recombination. We obtained *neo*+ colonies from cell lines with one or two *neo* genes, obtaining two- to threefold more colonies from the FN2 cell line than from the F, N2 or FN cell lines. The observed difference in the frequency of recombinants is consistent with the number of *neo* genes that can undergo DSB-promoted recombination (two for FN2; one each for F, N2 and FN), indicating that the presence of one or two DSBs does not significantly affect viability or the recovery of recombinants.

We used Southern blotting of genomic DNA from 43 *neo*+ clones derived from the FN2 cell lines to characterize the interchromosomal recombination products. A substantial portion (34 clones; 79% of total) had undergone interchromosomal recombination by gene conversion between the *neo* heteroalleles, without associated exchange of flanking restriction site markers (Fig. 2; data not shown). In 24 of these clones (56% of total), the conversion tract was limited to the *neo* homology region (short tract gene conversion, STGC), although in the other ten (23% of total), the conversion tract extended into non-homologous sequences immediately flanking the *neo* homologies (long tract gene conversion, LTGC). The lack of gross chromosomal rearrangements was verified by fluorescence *in situ* hybridization (FISH; Fig. 3, data not shown). We performed additional Southern blot and polymerase chain reaction (PCR) analyses to characterize the nature of the events further (data not shown). Of the 34 gene conversion clones, none had both alleles converted to *neo*+, and there was no bias in conversion of the alleles. Of the 24 STGCs, 12 occurred at the F allele and 12 at the N2 allele, and of the ten LTGCs, six occurred at the F allele and four at the N2

allele. In addition, the I-SceI site on the non-converted chromosome was intact in 21 of the 24 STGC and eight of the ten LTGC clones, with the remaining clones having undergone NHEJ at the site.

The remaining nine *neo*+ clones (21% of total) all exhibited reciprocal translocations originating within the *neo* gene segments. Seven of these events were identified on the basis of Southern blotting and PCR using primers that flanked the breakpoint junctions (Fig. 2; data not shown), with confirmation by FISH in all seven (Table 2; Fig. 3c, d). In each of these, the F/N2 hybrid allele was the size predicted for an intact *neo*+ gene flanked by the hygromycin (*hyg*) and hypoxanthine phosphoribosyltransferase (*HPRT*) markers, whereas the reciprocal N2/F allele varied in size. In the remaining two clones (clones 21 and 47), reciprocal translocation events were confirmed by FISH analysis (Table 2; data not shown); however, although a *neo*+ F/N2 allele was detected by PCR, Southern and FISH analyses, the reciprocal N2/F allele was only observed by FISH analysis. Presumably this indicated a deletion of at least 2.4 kilobases (kb), including the *neo* sequences and PCR primer binding sites. These results indicated that the structures of all nine translocations were analogous.

We used PCR and sequencing to characterize the breakpoint junctions of the translocations. The hybrid F/N2 allele, amplified with primers in the *hyg* and *HPRT* sequences flanking the *neo*+ gene, gave rise to the expected 3.3-kb product in all nine translocation clones (data not shown). This represented a deletion of 535 base pairs (bp) at the overlap of the two *neo* genes. We confirmed two of the nine F/N2 junctions by sequencing (clones 12 and 20). We amplified the reciprocal N2/F allele with primers in the *Rb* and *pim-1* sequences, and the product, detectable in seven of the nine translocation clones, was variable in size amongst the different clones (Fig. 2c). We did not obtain any of the 2.9-kb products that would have been produced from a gene conversion associated with a reciprocal exchange. Instead, the largest product we detected was 2.4 kb (clones 7, 12 and 20), indicating the joining of the broken ends from the N2 and F alleles (Fig. 2a, c). Sequence analysis at the N2/F breakpoint junction demonstrated NHEJ with minimal deletion from the ends (Table 2). Sequence analysis of the clones with smaller PCR products also confirmed NHEJ events with deletions in the range of 0.1 to 1.3 kb (Table 2). Overall, the N2/F breakpoint junctions of the seven sequenced clones were similar to translocation breakpoints sequenced from leukaemic cell lines²⁰ and to those obtained from NHEJ at a DSB within a single chromosome⁹⁻¹¹.

These experiments demonstrate that two DSBs are sufficient to cause chromosomal translocations in mammalian cells. As clones with translocations were recovered at a frequency of 1×10^{-4} per viable cell after induction of DSBs, but not from repair of one DSB ($< 5 \times 10^{-8}$ per viable cell), a second DSB increases the frequency of translocations at least 2,000-fold. Notably, each of the reciprocal translocations has a particular structure produced by non-conservative SSA (F/N2 allele) and NHEJ (N2/F allele). The frequent translocation events are due, at least in part, to selection for these

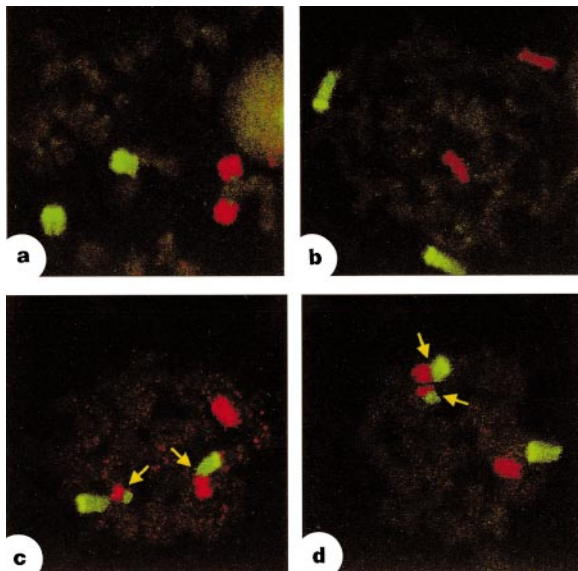


Figure 3 Visualization of translocations by FISH using whole chromosome mouse chromosome14-FITC (green) and chromosome17-Cy3 (red) probes. **a**, Parental FN2 and **b**, STGC clones have two normal chromosomes14 and 17. **c,d**, Reciprocal translocation clones 20 (**c**) and 12 (**d**) have normal untargeted chromosomes14 (green) and 17 (red), and monocentric hybrid chromosomes (red/green). Breakpoints are indicated by the arrows. Translocation chromosomes are approximately two-thirds (F/N2) and four-thirds (N2/F) of a chromosome length, as the I-SceI sites on chromosomes 17 and 14 were roughly one-third and two-thirds of the chromosome length from the centromere, respectively.

Table 2 Reciprocal translocations* induced by two DSBs

Clone number ¹	F/N2 junction	N2/F Junction	N2/F sequences
7	Δ535 bp†	Δ13 bp	CTTA-ATCC‡
12	Δ535 bp	Δ16 bp	TAAT-CAGGG
20	Δ535 bp	Δ26 bp	CAAT-CATG
5	Δ535 bp	Δ118 bp	TTAATAGC
27	Δ535 bp	Δ435 bp	GCAGGCTT
41	Δ535 bp	Δ1172 bp	CTTTAACA
8	Δ535 bp	Δ1276 bp	ACTGGTAG
21	Δ535 bp	>Δ2.4 kb	n.d.
47	Δ535 bp	>Δ2.4 kb	n.d.

* Reciprocal translocations were all verified by FISH.

† F/N2 junctions gave a net deletion of 535 bp at the sequence overlap due to SSA.

‡ N2/F junction sequences are 5' to 3' with the N2 and F sequences separated by a hyphen. In some cases, microhomology was present at the junction (underline). Clone 7 had a single base T insertion, and clone 20 had a 75-bp insertion at the junction. n.d., not determined.

alternative interchromosomal repair options resulting from a second DSB. The absence of non-reciprocal translocations indicates either that translocations are restricted to the G1 phase of the cell cycle or that a bias exists in the segregation of recombinant chromosomes during G2 events²³. In the latter case, recombinant chromosomes might always segregate to the same pole, consistent with the observed reciprocity, or to opposite poles, producing non-reciprocal translocations, and are not observed owing to imbalanced genetic information.

Although STGC and LTGC interchromosomal repairs events have been detected in the FN cell lines when only a single DSB was possible (97% and 3% of total, respectively)¹⁷, we observed a significantly higher frequency of both when two DSBs were possible in the FN2 cell lines. The 30-fold increase in repair frequency per allele, relative to the almost identical FN cell line, raises the possibility that, although the I-SceI site was retained on the non-converted allele in 85% of clones, the second chromosome in these clones was also cleaved but repaired to restore the I-SceI site. Precise repair could occur either by ligation of the compatible ends or by homologous sister chromatid repair during G2. The synergism could be due to such factors as assembly of broken chromosomes into the same repair complex or increased mobility of the ends, both of which could promote their interaction. Alternatively, an additional DSB may increase overall repair within the cell, although this was not evident in control gene targeting experiments (Table 1). The failure to recover clones in which both alleles had undergone gene conversion may indicate that two gene conversion events cannot be efficiently coupled.

Although translocation junctions in leukaemic cells often appear to be classic NHEJ junctions²⁰, chromosome rearrangements at or near repetitive elements, especially Alu elements, have also been reported. Events involving repetitive elements are associated with genetic diseases¹⁶ and tumorigenesis, as a result of rearrangements including gene deletions and duplications²⁴, "jumping" translocations²⁵ and proto-oncogene amplification²⁶. The results presented here suggest that such rearrangements, although promoted by homology, may in some instances arise from non-conservative SSA repair. Thus, cells in which gene conversion is perturbed could be at high risk for chromosome rearrangements owing to a reliance on non-conservative repair mechanisms. Supporting this, cells mutated in the hereditary breast cancer gene *BRCA1* are defective in homologous repair of DSBs²⁷ and they exhibit a high level of aberrations²⁸. The system we have developed to study translocations will allow a determination of the genetic requirements for the suppression of these events. □

Methods

DNA manipulations and cell transfections

The N2 targeting vector was adapted from the N targeting vector³⁹ by *PacI* digestion, T4 polymerase treatment, and ligation to a double-stranded oligonucleotide containing the I-SceI recognition site 5'-ATTACCCTGTTATCCCTATG-3'. F cells¹⁷ were targeted by electroporation of N2 vector DNA to create the FN2 cell lines. Mouse E14TG2a ES cells (WT) were electroporated with one (F or N2) or both constructs (FN2) and selected in hygromycin or HAT (0.1 mM hypoxanthine, 0.8 μM aminopterin, 20 μM thymidine)¹⁷. Targeted clones were identified by diagnostic digests and Southern blot analysis with probes from genomic sequences outside the constructs as described¹⁷. To examine DSB-induced recombination, 50 μg of the circular pCBASce expression vector or 40 μg of circular pneo-WT was electroporated with 0.8 ml cells at 2×10^7 or 1×10^8 cells ml⁻¹. After 20 h, G418 was added at 200 μg ml⁻¹. After selection, independent G418 resistant colonies were counted and expanded. The donor plasmid pneo-WT contains an internal *neo* gene fragment with homology that flanks the insertion mutations in the *neo* genes¹⁷.

DNA analysis

DNA extractions and Southern blot analysis were performed as described¹⁷. In addition to the Southern analysis in Fig. 2, further Southern and PCR analyses were performed using *NotI* and I-SceI endonuclease digests, respectively. PCR for amplification of repair junctions was performed to amplify the following with the specified primers: the hybrid F/N2 allele, *hyg* (5'-TACTCGCCGATAGTGGAAC-3') and *HPRT* (5'-ACAAGTTTGTGTGGATAT GCC-3'); the hybrid N2/F allele, exon 20 of *Rb* (5'-GTATGCTGAAACAGAGTCC-3') and intron 4 of *pim-1* (5'-CCAATCACGCCA-GAAATGTT-3'); and an internal *neo* fragment (5'-GACAATCGGCTGCTCTGATG-3'

and 5'-CTCGCTCGATGCGATGTTTC-3'). Amplification was performed by denaturation at 94 °C for 5 min; followed by 40 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 2 min; and extension at 72 °C for 15 min. PCR products were cloned with the TA cloning system (Invitrogen), and sequenced by the Sloan-Kettering Institute core facility.

FISH

Chromosome metaphase spreads from individual clones were made as described¹⁷. Spreads were hybridized to whole chromosome mouse chromosome14-FITC and chromosome17-Cy3 probes (Vysis-Cambio). Chromosomes were visualized by confocal microscopy at the Sloan-Kettering Institute core facility. All LTGC and reciprocal translocation clones and four STGC clones were analysed by FISH.

Received 20 September 1999; accepted 17 March 2000.

- Rabbitts, T. H. Chromosomal translocations in human cancer. *Nature* **372**, 143–149 (1994).
- Mitelman, F., Mertens, F. & Johansson, B. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genet.* **15**, 417–474 (1997).
- Tycko, B. & Sklar, J. Chromosomal translocations in lymphoid neoplasia: a reappraisal of the recombinase model. *Cancer Cells* **2**, 1–8 (1990).
- Lewis, S. M. The mechanism of V(D)J joining: lessons from molecular, immunological and comparative analyses. *Adv. Immunol.* **56**, 27–149 (1994).
- Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
- Cornforth, M. N. & Bedford, J. S. Ionizing radiation damage and its early development in chromosomes. *Adv. Radiat. Biol.* **17**, 423–496 (1993).
- Ikedo, H. DNA topoisomerase-mediated illegitimate recombination. *Adv. Pharmacol.* **29A**, 147–165 (1994).
- Wang, P., Zhou, R., Zou, Y., Jackson-Cook, C. & Povirk, L. Highly conservative reciprocal translocations formed by apparent joining of exchanged DNA double-strand break ends. *Proc. Natl Acad. Sci. USA* **94**, 12018–12023 (1997).
- Rouet, P., Smih, E. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
- Sargent, R. G., Breneman, M. A. & Wilson, J. H. Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination. *Mol. Cell. Biol.* **17**, 267–277 (1997).
- Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 5172–5177 (1998).
- Kadyk, L. C. & Hartwell, L. H. Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics* **132**, 387–402 (1992).
- Johnson, R. D., Liu, N. & Jasin, M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. *Nature* **401**, 397–399 (1999).
- Lin, F. L., Sperle, K. & Sternberg, N. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. *Mol. Cell. Biol.* **4**, 1020–1034 (1984).
- Jeggo, P. A. DNA breakage and repair. *Adv. Genet.* **38**, 185–218 (1998).
- Cooper, D. N., Krawczak, M. & Antonarakis, S. E. In *The Genetic Basis of Human Cancer* (eds Vogelstein, B. & Kinzler, K. W.) 65–94 (McGraw-Hill, New York, 1998).
- Richardson, C., Moynahan, M. E. & Jasin, M. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev.* **12**, 3831–3842 (1998).
- Haber, J. E. & Leung, W. Y. Lack of chromosome territoriality in yeast: promiscuous rejoining of broken chromosome ends. *Proc. Natl Acad. Sci. USA* **93**, 13949–13954 (1996).
- Chen, C. & Kolodner, R. D. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nature Genet.* **23**, 8–85 (1999).
- Gillert, E. *et al.* A DNA damage repair mechanism is involved in the origin of chromosomal translocations (t(4;11) in primary leukemic cells. *Oncogene* **18**, 4663–4671 (1999).
- Colleaux, L., d'Auriol, L., Gailbert, F. & Dujon, B. Recognition and cleavage site of the intron-encoded *omega* transposase. *Proc. Natl Acad. Sci.* **85**, 6022–6026 (1988).
- Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**, 224–228 (1996).
- Beumer, K. J., Pimpinelli, S. & Golic, K. G. Induced chromosomal exchange directs the segregation of recombinant chromatids in mitosis of *Drosophila*. *Genetics* **150**, 173–188 (1998).
- Strout, M. P., Marcucci, G., Bloomfield, C. D. & Caligiuri, M. A. The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. *Proc. Natl Acad. Sci. USA* **95**, 2390–2395 (1998).
- Andreasson, P. *et al.* Molecular characterization of jumping translocations reveals spatial and temporal breakpoint heterogeneity. *Leukemia* **12**, 1411–1416 (1998).
- Xu, X. *et al.* Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in *BRCA1* exon 11 isoform-deficient cells. *Mol. Cell* **3**, 389–395 (1999).
- Coleman, A. E., Kovalchuk, A. L., Janz, S., Palini, A. & Ried, T. Jumping translocation breakpoint regions lead to amplification of rearranged myc. *Blood* **93**, 4442–4444 (1999).
- Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. *Brc1* controls homology-directed repair. *Mol. Cell* **4**, 511–518 (1999).
- Moynahan, M. E. & Jasin, M. Loss of heterozygosity induced by a chromosomal double-strand break. *Proc. Natl Acad. Sci. USA* **94**, 8988–8993 (1997).

Acknowledgements

We thank Hein te Riele (Amsterdam) for materials, D. Tabarini in the core sequencing facility, and K. Manova and S. Kerns in the core microscopy facility. C.R. is a recipient of the Vrushalli Ranadive Fellowship from the Leukemia and Lymphoma Society, formerly the Leukemia Society of America. This work was supported by a grant from the National Science Foundation to M.J.

Correspondence and requests for materials should be addressed to M.J. (e-mail: m-jasin@ski.mskcc.org).

Redundant roles for the TFIID and SAGA complexes in global transcription

Tong Ihn Lee^{*†‡}, Helen C. Causton^{*‡}, Frank C. P. Holstege^{*§}, Wu-Cheng Shen^{||}, Nancy Hannett^{*}, Ezra G. Jennings^{*†}, Fred Winston[¶], Michael R. Green^{||} & Richard A. Young^{*†}

^{*} Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

[†] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[§] Department for Medical Genetics, University Medical Centre Utrecht, Lunlaan 6, 3584 EA Utrecht, The Netherlands

^{||} Howard Hughes Medical Institute Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, Massachusetts 01605, USA

[¶] Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

[‡] These authors contributed equally to this work

The transcription factors TFIID and SAGA are multi-subunit complexes involved in transcription by RNA polymerase II^{1,2}. TFIID and SAGA contain common TATA-binding protein (TBP)-associated factor (TAF_{II}) subunits and each complex contains a subunit with histone acetyltransferase activity³. These observations have raised questions about whether the functions of the two complexes *in vivo* are unique or overlapping. Here we use genome-wide expression analysis to investigate how expression of the yeast genome depends on both shared and unique subunits of these two complexes. We find that expression of most genes requires one or more of the common TAF_{II} subunits, indicating that the functions of TFIID and SAGA are widely required for gene expression. Among the subunits shared by TFIID and SAGA are three histone-like TAF_{II}s, which have been proposed to form a sub-complex and mediate a common function in global transcription. Unexpectedly, we find that the histone-like TAF_{II}s have distinct roles in expression of the yeast genome. Most importantly, we show that the histone acetylase components of TFIID and SAGA (TAF_{II}145 and Gcn5) are functionally redundant, indicating that expression of a large fraction of yeast genes can be regulated through the action of either complex.

To assess the requirement for components of the TFIID and SAGA transcriptional co-activator complexes *in vivo*, we used high-

density oligonucleotide arrays (HDAs)⁴ to determine the genome-wide effects on transcription caused by inactivation or loss of components of these two complexes. We used conditional lethal mutations in genes encoding essential components of TFIID and SAGA and complete deletions of genes encoding nonessential components. Understanding of the contributions of TAF_{II} subunits has been complicated by the substantially different experimental and analytical approaches used in previous *in vivo* studies^{3,5–16}; here we have applied a consistent and systematic approach to the analysis of TAF_{II} function. Additional scientific information on TFIID and SAGA, details of the experimental reagents and approaches, interactive databases supporting this work and more extensive discussion of the results can be found at our website¹⁷. Raw data is available as Supplementary Information.

We used HDA analysis to assess whether subunits of TFIID and SAGA have global roles in transcription. Temperature-sensitive mutations in all five shared subunits of TFIID and SAGA (TAF_{II}90, TAF_{II}61/68, TAF_{II}60, TAF_{II}25 and TAF_{II}17) were used to identify the sets of genes whose expression depends on these components; the results for TAF_{II}17 have been reported^{10,11}. The effects of inactivating individual TAF_{II}s are summarized in Table 1. Although four shared TAF_{II} subunits have been reported to have general functions^{12,14,15}, no single TAF_{II} examined here was found to be required to the same extent as RNA polymerase II (*rpb1-1* fit criteria)¹¹. Instead, each shared TAF_{II} subunit was necessary for expression of a characteristic subset of genes, ranging from 8% to 59% of the genome. Upon identification of genes whose expression depends on one or more of these five TAF_{II}s, we find that shared TAF_{II}s are required for normal expression of 70% of the genome (Fig. 1a). If these shared TAF_{II} subunits are components of SAGA and TFIID, but not other complexes, the results indicate that SAGA and TFIID are involved in the expression of around 70% of yeast genes.

We considered the possibility that the different signatures seen with individual subunits might reflect differences in the strengths of the various temperature-sensitive alleles rather than different requirements for specific subunits. Although we cannot eliminate

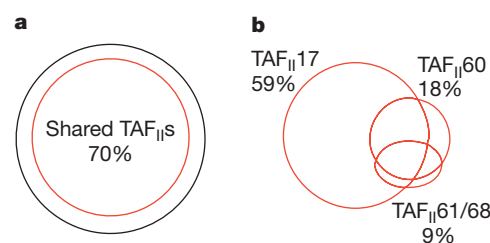


Figure 1 Whole-genome analysis of genes affected by the shared TAF_{II}s. **a**, The sum of individual subunit contributions reveals a global role for shared TAF_{II}s. Venn diagram showing genes requiring one or more of the shared TAF_{II}s for normal expression as a percentage of genes comparable across these datasets. **b**, Genes affected by histone-like TAF_{II}s overlap significantly, although each of the histone-like TAF_{II}s affect genome-wide transcription to different extents.

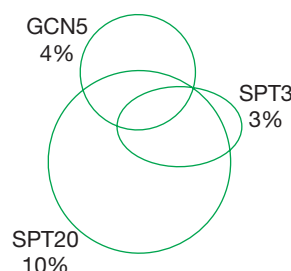


Figure 2 Expression of distinct sets of genes depends on individual subunits of SAGA. Venn diagram describing the overlap of genes affected by deletions of *SPT3*, *GCN5* or *SPT20*.

Table 1 Percentage of genome dependent on subunits of SAGA and TFIID

Subunit		% Genome dependent	
		2-fold down	rpb-1 fit criteria
Shared	TAF _{II} 17	67%	59%
	TAF _{II} 25	38%	16%
	TAF _{II} 60	34%	18%
	TAF _{II} 61/68	11%	9%
	TAF _{II} 90	10%	8%
TFIID-specific	TAF _{II} 145	27%	14%
	TAF _{II} 150	2%	3%
SAGA-specific	SPT3	3%	–
	SPT20	10%	–
	GCN5	4%	–

Two approaches were used to compare datasets, as described previously¹¹. We report genes that showed a 2-fold or greater decrease in mRNA levels in two independent HDA experiments for all mutant strains. For strains with conditional mutations, expression levels are also compared to a dataset generated with an *rpb1-1* ts allele. This comparison identifies genes whose expression is as dependent on the factor of interest as it is on core RNA polymerase II. Consequently, this comparison provides a description of genes which are most likely to be directly affected by the loss of this factor. For strains with deletion mutations, where no such comparison can be made, the genes affected necessarily reflect both direct and indirect effects of loss of these factors. For lists of genes whose expression levels decrease or increase upon loss of SAGA or TFIID subunits, see ref. 17 or Supplementary Information.

this possibility, we consider it unlikely for several reasons. First, for every TAF_{II} dataset there is a set of genes for which the loss of expression upon TAF_{II} inactivation parallels the loss of expression upon RNA polymerase II inactivation by the *rpb1-1* temperature-sensitive allele. This indicates that the TAF_{II} inactivation must have been as rapid as polymerase inactivation by the *rpb1-1* allele, which occurs within five minutes¹¹. Second, different temperature-sensitive alleles of any one TAF_{II} produce highly similar transcription profiles, as 80–85% of the genes affected by one allele are affected by a second allele (data not shown). This observation indicates that the specific signatures obtained for different TAF_{II}s are not simply a consequence of allele-specific inactivation of each TAF_{II}. Third, growth rates at the restrictive temperature, commonly used to estimate the strength of temperature-sensitive alleles, are essentially identical in cells harbouring mutations in TAF_{II}145, TAF_{II}61/68, TAF_{II}60 and TAF_{II}17 (ref. 17). Finally, evidence that TAF_{II}145 alone has a histone acetyltransferase activity¹⁸ indicates that there are functional differences between TFIID subunits that should be reflected in differential requirements for specific subunits.

Temperature-sensitive mutations in one of the TAF_{II}s shared by TFIID and SAGA, TAF_{II}90, cause cells to arrest growth at the G₂/M stage of the cell cycle⁹. We examined the expression data for effects on genes that might explain the arrest phenotype and identified two candidate genes: *SPC98*, a component of the spindle pole body, and *APC2*, a component of the anaphase-promoting complex. Loss-of-function mutations in these genes result in strains that arrest with large budded cells, consistent with a G₂/M cell-cycle arrest.

Three shared TAF_{II}s (TAF_{II}17, TAF_{II}60 and TAF_{II}61/68) interact through conserved protein–protein interaction motifs and have been proposed to form a nucleosome-like structure involved in contacts between TFIID and DNA^{19–21}. The loss of these TAF_{II}s

affects genome-wide expression to different extents, as mutations in TAF_{II}17, TAF_{II}60 and TAF_{II}61/68 affect 59%, 18% and 9% of the genome, respectively (Table 1). The genes affected by TAF_{II}17, TAF_{II}60 and TAF_{II}61/68 overlap significantly (Fig. 1b), indicating that these TAF_{II}s function together at many promoters, but differences in their genome-wide requirements indicate that the individual subunits must make distinct contributions to transcription *in vivo*.

To explore further the extent of TFIID and SAGA function *in vivo*, cells with mutations in TFIID-specific subunits (TAF_{II}150 and TAF_{II}145) and SAGA-specific subunits (Spt3, Spt20 and Gcn5) were subjected to genome-wide expression analysis. The results (Table 1) indicate that each mutant affects a distinct and generally small subset of genes and are most consistent with the idea that subunits of TFIID and SAGA make specific and unique contributions to the functions of the complete complex. Expression of no more than one-eighth of the genome is affected by the loss of Gcn5, Spt3 and Spt20 activity. When assessed by the same criteria used to evaluate the effects of mutations in SAGA components, mutations in the TFIID-specific subunits TAF_{II}150 and TAF_{II}145 together affect expression of less than one-third of the genome.

Spt3, Spt20 and Gcn5 were chosen as representatives of each of three subcomplexes within SAGA that have been defined previously^{22–27}. The sets of genes affected by the loss of these three components (Fig. 2) are consistent with previous biochemical and genetic evidence indicating that Spt3 and Gcn5 control distinct sets of genes, and that Spt20, which is required for the integrity of SAGA, controls a larger set of genes. The set of genes affected by loss of Spt3 is mostly included within the larger set of Spt20-dependent genes. The set of genes affected by loss of Gcn5 contains both Spt20-dependent and Spt20-independent genes, consistent with the finding that Gcn5 exists in additional complexes^{22,26,27}.

As many SAGA components were originally identified by their effects on genes whose expression was influenced by transposon promoters²⁸, we examined the expression data for genes affected by SAGA disruption. Of the 30 transposon open reading frames (ORFs) represented on the HDAs, all 24 for which a signal could be detected were dependent on one or more SAGA subunits (Table 2). In addition, SAGA may co-regulate an interface between phosphate metabolism and cell-cycle control, as genes involved in

Table 2 Genes affected by disruption of SAGA

Gene	Description
Transposon elements	
YAR009C	Ty element
YBR012WA	Ty element
YER138C	Ty element
YER160C	Ty element
YMR050C	Ty element
YAR010C	Ty element
YBL005WA	Ty element
YBL005WB	Ty element
YBL101WA	Ty element
YBL101WB	Ty element
YBR012WB	Ty element
YCL019W	Ty element
YCL020W	Ty element
YHR214CB	Ty element
YJR026W	Ty element
YJR027W	Ty element
YJR028W	Ty element
YJR029W	Ty element
YML039W	Ty element
YML040W	Ty element
YML045W	Ty element
YMR045C	Ty element
YMR046C	Ty element
YMR051C	Ty element
Phosphate uptake	
PHO3	Acid phosphatase, constitutive
PHO5	Acid phosphatase, repressible
PHO11	Acid phosphatase, inducible
PHO12	Acid phosphatase, inducible
PHO84	High affinity inorganic phosphate/H ⁺ symporter
Pho80–Pho85 function	
PHO81	CDK inhibitor of Pho80–Pho85
SPL2	Putative inhibitor of Pho80–Pho85 CDK complex
YJL012C	Similar to PHO81, YPL019C
YPL019C	Similar to YJL012C, NRF1
NRF1	Negative regulator of CDC42
YDR281C	Induced by an inhibitor of Pho85

Of 30 transposon ORFs represented on the HDAs, all 24 for which a signal could be obtained were dependent on one or more of the SAGA components studied here. Among the set of genes which are dependent on Gcn5, Spt3 and Spt20 are several genes involved in phosphate uptake and known and potential regulators of the Pho80–Pho85 cyclin-dependent kinase complex.

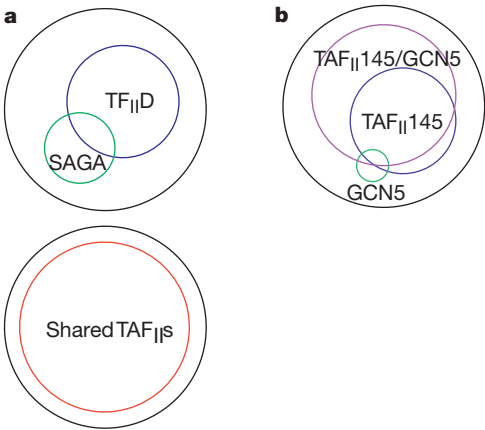


Figure 3 TFIID and SAGA have compensatory functions at a large fraction of genes. **a**, Venn diagrams showing genes dependent on TFIID-specific and SAGA-specific subunits compared to the set of genes dependent on subunits shared between the two complexes. The complete genome is represented by the black circle. A minimum of ~30% of the genome is dependent on TFIID-specific subunits whereas ~12% of the genome is dependent on SAGA-specific subunits. In contrast, ~70% of the genome depends on one or more of the shared TAF_{II}s. **b**, Venn diagram showing the overlap of genes affected by the *TAF_{II}145* mutation, the *gcn5* mutation and the *TAF_{II}145/gcn5* double mutation. The complete genome is represented by the black circle. The expression of ~25% of yeast genes decreases 2-fold or more only when both *TAF_{II}145* and *Gcn5* are depleted.

uptake of extracellular phosphate and regulators of the Pho80–Pho85 cyclin-dependent kinase complex are dependent on SAGA components (Table 2).

Genome-wide expression analysis of the role of TAF_{II}s shared between TFIID and SAGA shows that some combination of TFIID and SAGA function is probably required for transcription of about 70% of the yeast genome. In contrast, the sum of effects seen with mutations in complex-specific subunits was substantially less than the sum of effects of the shared TAF_{II}s (Fig. 3a). HDA analysis of additional complex-specific subunits may reveal further effects of TFIID and SAGA which could account for the nearly global requirement for the shared TAF_{II}s. Alternatively, TFIID and SAGA might have compensatory activities, such that the loss of one complex could be partially rescued by redundant functions of the other.

As both TFIID and SAGA have subunits with histone acetyltransferase activity, we tested the possibility that SAGA and TFIID have compensatory histone acetylase components. We generated a double *TAF_{II}145/gcn5* mutant by deleting *GCN5* in the *TAF_{II}145* temperature-sensitive background. Genome-wide expression data from this mutant was compared to that obtained from the single *TAF_{II}145* and *gcn5* mutations. The results reveal that for around one-quarter of the genome, changes in expression are only apparent

when the functions of both TAF_{II}145 and Gcn5 are lost (Fig. 3b) and identify a set of genes that appear to require either TAF_{II}145 or Gcn5 for normal expression (Table 3).

Two conclusions follow from these observations. First, there are distinct requirements for specific subunits of TFIID and SAGA in global expression. Second, the functions of the TAF_{II}145 and Gcn5 subunits, and possibly the protein complexes with which they are associated, are redundant. By redundant, we mean that the function of one subunit can compensate for the loss of the other, although the precise mechanism remains to be determined. Gcn5 is also redundant with the Swi/Snf nucleosome remodelling complex^{24,29,30}, highlighting the complex interactions between large transcription complexes *in vivo*. Further genome-wide expression analysis should help to reveal the requirements for all transcription components in the global regulatory network. □

Methods

Strains, media and growth conditions

The *TAF_{II} 60* and *TAF_{II} 61/68* alleles used here have been described¹². The *TAF_{II} 145*, and *TAF_{II} 90* alleles were previously described and *TAF_{II} 150* and *TAF_{II} 25* alleles were generated as described^{6,9}. The *spt3*, *spt20* and *gcn5* alleles have been described²⁴. Additional information regarding these strains can be found at the accompanying website¹⁷.

For genome-wide expression experiments, two wild-type and two mutant cultures were grown to an OD₆₀₀ of 0.5–0.8 in YPD at 30 °C. Cells were then collected if the experiment involved strains with deletion mutations. For strains with temperature-sensitive mutations, all four cultures were rapidly shifted to the nonpermissive temperature by addition of an equal volume of YPD at 44 °C and allowed to grow for an additional 45 min at 37 °C. Total RNA and then polyA+ RNA was isolated and used to generate complementary DNA and biotin–cRNA, as described¹¹.

Genome-wide expression profiling

Biotin-labelled antisense RNA target was prepared as described¹¹. The cRNA was hybridized to a set of four oligonucleotide arrays (GeneChip Ye6100 arrays, Affymetrix) and scanned as described⁴. Intensities were captured using GeneChip software (Affymetrix) and a single raw expression level for each gene was determined.

Data analysis

The raw data from each chip was normalized and corrected as described¹¹. A change in messenger RNA levels was deemed significant and reported in the lists containing genes >2-fold down (or up) on the basis of the following criteria: the average fold change was more than 2-fold, the fold change was consistent in duplicate experiments, and the change in the values was above background values in duplicate experiments. Comparisons of 2-fold-down lists across multiple datasets were performed with those genes which could be consistently scored across all datasets.

Received 9 November 1999; accepted 10 March 2000.

- Hahn, S. The role of TAFs in RNA polymerase II transcription. *Cell* **95**, 579–582 (1998).
- Lee, T. & Young, R. Regulation of gene expression by TBP-associated proteins. *Genes Dev.* **12**, 1398–1408 (1998).
- Grant, P. *et al.* A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* **94**, 45–53 (1998).
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H. & Lockhart, D. J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol.* **15**, 1359–1367 (1997).
- Moqtaderi, Z., Bai, Y., Poon, D., Weil, P. A. & Struhl, K. TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* **383**, 188–191 (1996).
- Walker, S. S., Reese, J. C., Apone, L. M. & Green, M. R. Transcription activation in cells lacking TAFII. *Nature* **383**, 185–188 (1996).
- Walker, S. S., Shen, W. C., Reese, J. C., Apone, L. M. & Green, M. R. Yeast TAF(II)145 required for transcription of G1/S cyclin genes and regulated by the cellular growth state. *Cell* **90**, 607–614 (1997).
- Shen, W. & Green, M. Yeast TAF(II)145 functions as a core promoter selectivity factor, not a general coactivator. *Cell* **90**, 615–624 (1997).
- Apone, L. M., Virbasius, C. -M. A., Reese, J. C. & Green, M. R. Yeast TAF(II)90 is required for cell-cycle progression through G2/M but not for general transcription activation. *Genes Dev.* **10**, 2368–2380 (1996).
- Apone, L. *et al.* Broad, but not universal, transcriptional requirement for yTAFII17, a histone H3-like TAFII present in TFIID and SAGA. *Mol. Cell* **2**, 653–661 (1998).
- Holstege, F. C. P. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
- Michel, B., Komarnitsky, P. & Buratowski, S. Histone-like TAFs are essential for transcription *in vivo*. *Mol. Cell* **2**, 663–673 (1998).
- Moqtaderi, Z., Keaveney, M. & Struhl, K. The histone H3-like TAF is broadly required for transcription in yeast. *Mol. Cell* **2**, 675–682 (1998).
- Natarajan, K., Jackson, B., Rhee, E. & Hinnebusch, A. yTAFII61 has a general role in RNA polymerase II transcription and is required by Gcn4p to recruit the SAGA coactivator complex. *Mol. Cell* **2**, 683–692 (1998).
- Sanders, S. L., Klebanow, E. R. & Weil, P. A. TAF25p, a non-histone-like subunit of TFIID and SAGA complexes, is essential for total mRNA gene transcription *in vivo*. *J. Biol. Chem.* **274**, 18847–18850 (1999).

Table 3 Selected genes with differential requirements for TFIID and SAGA

TFIID-dependent*	
MSH6	Involved in repair of single base mismatches
DBP1	Similar to DEAD/DEAH box RNA helicases
KEM1	Nuclease, degrades decapped mRNA
WSC2	Involved in cell wall integrity and stress response
TRX1	Thioredoxin I
GNA1	Putative acetyltransferase
HTA3	Histone-related protein
WTM2	Transcription modulator of meiosis and silencing
CAM1	Translation elongation factor
PMT4	Mannosyltransferase
SAGA-dependent†	
PHO84	High-affinity inorganic phosphate H ⁺ symporter
SPL2	Putative inhibitor of Pho80–Pho85 CDK complex
YHB1	Flavohemoglobin of unknown function
YPL019C	Similar to YJL012C, NRF1
PHO5	Acid phosphatase, repressible
PFK27	6-phosphofructose-2-kinase, isozyme 2
YDL038C	Protein of unknown function
NRF1	Protein of unknown function
YDL124W	Protein of unknown function
YJL012C	Similar to PHO81, YPL019C
Dependent on both TFIID and SAGA‡	
CLB1	G2/M cyclin
UTR2	Protein of unknown function
YHR032C	Member of major facilitator superfamily
AAH1	Adenine deaminase
PHO12	Acid phosphatase, secreted
PHO3	Acid phosphatase constitutive
PHO81	Inhibitor of Pho80–Pho85 CDK complex
FAR1	Inhibitor of CDC28 kinase complexes
YOR095C	Ribose 5-phosphate ketol-isomerase
YOR06C	Similar to human X-linked PEST-containing transporter
Dependent on either TFIID or SAGA§	
PAM1	Suppressor of loss of PP2A
YNR059W	Similar to MNN1
MBP1	Regulates MluI cell cycle box genes
YNL177C	Protein of unknown function
RIM2	Required for respiration
YDL177C	Protein of unknown function
HEM4	Uroporphyrinogen III synthase
TOM7	Translocase involved in mitochondrial import
MRP2	Mitochondrial ribosomal protein
PET127	Mitochondrial membrane-associated protein

Genes were scored using a 2-fold cutoff in duplicate experiments. Selection criteria included high expression level, comparability across all datasets and consistency of effect across all experiments.

* TFIID dependent: genes scored in TAF_{II}145 or TAF_{II}150 and not in SPT3, GCN5 or SPT20.

† SAGA dependent: genes scored in two of SPT3, GCN5 or SPT20 and not scored in TAF_{II}145 or TAF_{II}150.

‡ Dependent on both TFIID and SAGA: genes scored in TAF_{II}145 or TAF_{II}150 that also scored in SPT3, GCN5 or SPT20.

§ Dependent on either TFIID or SAGA: genes scored in the TAF_{II}145/GCN5 double mutant and not in TAF_{II}145, TAF_{II}150, SPT3, GCN5 or SPT20.

16. Komarnitsky, P. B., Michel, B. & Buratowski, S. TFIID-specific yeast TAF40 is essential for the majority of RNA polymerase II-mediated transcription in vivo. *Genes Dev.* **13**, 2484–2489 (1999).
17. Website accompanying this paper. http://web.wi.mit.edu/young/TFIID_SAGA
18. Mizzen, C. A. *et al.* The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261–1270 (1996).
19. Hoffmann, A. *et al.* A histone octamer-like structure within TFIID. *Nature* **380**, 356–359 (1996).
20. Xie, X. *et al.* Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**, 316–322 (1996).
21. Oelgeschlaeger, T., Chiang, C. M. & Roeder, R. G. Topology and reorganization of a human TFIID-promoter complex. *Nature* **382**, 735–738 (1996).
22. Horiuchi, J., Silverman, N., Pina, B., Marcus, G. & Guarente, L. ADA1, a novel component of the ADA/GCN5 complex, has broader effects than GCN5, ADA2, or ADA3. *Mol. Cell. Biol.* **17**, 3220–3228 (1997).
23. Roberts, S. M. & Winston, F. SPT20/ADA5 encodes a novel protein functionally related to the TATA-binding protein and important for transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 3206–3213 (1996).
24. Roberts, S. M. & Winston, F. Essential functional interactions of SAGA, a *Saccharomyces cerevisiae* complex of Spt, Ada, and Gcn5 proteins, with the Snf/Swi and Srb/mediator complexes. *Genetics* **147**, 451–465 (1997).
25. Sterner, D. *et al.* Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. *Mol. Cell. Biol.* **19**, 86–98 (1999).
26. Grant, P. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* **11**, 1640–1650 (1997).
27. Eberhart, A. *et al.* The ADA complex is a distinct histone acetyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 6621–6631 (1999).
28. Winston, F. & Sudarsanam, P. The SAGA of Spt proteins and transcriptional analysis in yeast: past, present, and future. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 553–561 (1998).
29. Sudarsanam, P., Cao, Y., Wu, L., Laurent, B. & Winston, F. The nucleosome remodeling complex, Snf/Swi, is required for the maintenance of transcription *in vivo* and is partially redundant with the histone acetyltransferase, Gcn5. *EMBO J.* **18**, 3101–3106 (1999).
30. Biggar, S. & Crabtree, G. Continuous and widespread roles for the Swi-Snf complex in transcription. *EMBO J.* **18**, 2254–2264 (1999).

Supplementary information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Workman for helpful discussions. This work was supported by funds from the NIH. F.C.P.H. was supported by a fellowship from the Human Frontier Science Program and E.G.J. is a predoctoral fellow of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.A.Y. (e-mail: young@wi.mit.edu).

Effects of mechanical forces on maintenance and adaptation of form in trabecular bone

Rik Huiskes*, Ronald Ruimerman*†, G. Harry van Lenthe* & Jan D. Janssen‡

* Orthopaedic Research Lab, Department Orthopaedics, University of Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

‡ Faculty of Biomedical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

The architecture of trabecular bone, the porous bone found in the spine and at articulating joints, provides the requirements for optimal load transfer, by pairing suitable strength and stiffness to minimal weight according to rules of mathematical design^{1–6}. But, as it is unlikely that the architecture is fully pre-programmed in the genes⁷, how are the bone cells informed about these rules, which so obviously dictate architecture? A relationship exists between bone architecture and mechanical usage⁸—while strenuous exercise increases bone mass⁹, disuse, as in microgravity and

inactivity, reduces it¹⁰. Bone resorption cells (osteoclasts) and bone formation cells (osteoblasts) normally balance bone mass in a coupled homeostatic process of remodelling, which renews some 25% of trabecular bone volume per year. Here we present a computational model of the metabolic process in bone that confirms that cell coupling is governed by feedback from mechanical load transfer^{11–13}. This model can explain the emergence and maintenance of trabecular architecture as an optimal mechanical structure, as well as its adaptation to alternative external loads.

Investigations have indicated that osteocytes may have a prominent function in the transduction of mechanical signals^{6,8,14–16} (Fig. 1). Osteocytes derive from osteoblasts which are entrapped in their own matrix and survive there, mutually connected by a network of canaliculi. This network also connects with the lining cells, which also derive from osteoblasts, that cover the trabecular surface. Hence, osteoblasts, osteocytes and lining cells form a syncytium, which is well-equipped for signal transduction. It is possible that increased strain in the local mineralized matrix signals the osteocyte to transmit stimuli to the surface, where bone is formed until the strains are normalized¹⁴. This paradigm provides an explanation for the mechanical control (growth and adaptation) of bone modelling¹¹.

Bone remodelling (or bone maintenance) is the formation of cavities by osteoclasts, and the subsequent filling of these cavities by osteoblasts (Fig. 2). We assume that the coupling factor between osteoblasts and osteoclasts in this process, as in bone modelling, is mechanical^{11–13}. While higher external forces increase the strains in the mineralized matrix of bone at large, the resorption cavities produced in bone remodelling have a similar strain-enhancing effect locally¹⁷ (Fig. 2). Thus, modelling and remodelling could both be described as being governed by strain perturbations, be they generated externally by the load or internally by resorption cavities. Using refined finite-element methods of stress analysis (μ -FEA)¹⁸ in a regulatory computer-simulation model, we have evaluated these local strain perturbations and determined whether this unified mechanical paradigm can indeed explain the morphological phenomena in bone.

The regulatory process that we propose is depicted in Fig. 3 and involves the following hypotheses. (1) The mechanical variable that triggers feed-back from the external forces to bone metabolism is a typical strain-energy density (SED) rate in the mineralized tissue, as produced by a recent loading history. Loading rate has been indicated as a bone-producing factor in a rat model¹⁹. Strain-energy density can be linked to trabecular architecture as an optimal structure^{5,6}. Both the frequency and the amplitude of the external loads have a role²⁰, and a limited amount of loading cycles per day suffice for the maintenance of bone mass²¹. All of this information is incorporated in a typical, daily SED rate as a strain-derived scalar signal.

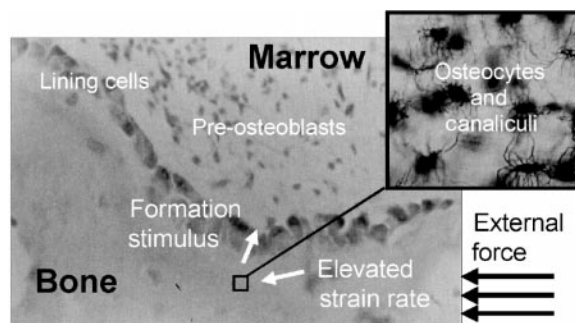


Figure 1 Osteocytes are probably mechanosensors which send strain-related signals to lining cells located at the bone surface through the canalicular syncytium. These stimuli are thought to recruit osteoblasts, which add bone to the surface when strains increase¹⁴.

† Present address: Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(2) Osteocytes react to the loading in their local environments by producing a biochemical messenger in proportion to the typical SED rate. Osteocytes are well placed to function as 'strain gauges', and no other function for them has been suggested⁸. Osteocytes react quickly to mechanical stimulation *in vivo*¹⁵. They are extremely sensitive to fluid shear stresses *in vitro*, producing biochemical messengers such as prostaglandin²². The SED rate can potentially provide the pumping action to enforce fluid flow through the canaliculi and over the cell membranes^{14,23}.

(3) The biochemical messenger produced by the osteocytes causes signals to be dissipated through the osteocytic network towards the bone surface, where they create an osteoblast recruitment stimulus¹⁴ (Fig. 1). The strength of this signal, which is produced by all osteocytes in the environment, stimulates osteoblast recruitment and bone formation as long as it exceeds a threshold value. The nature of such a signal is unknown, although both electric current and ion transport have been suggested^{14,23}. In our model, we link the bone formation stimulus at the surface directly to the SED rate per osteocyte, subject to osteocyte mechanosensitivity and attenuation by distance.

(4) To portray resorption in the model, the probability p of osteoclast activation per surface site at any time is considered to be regulated either by the presence of microcracks within the bone matrix (hypothesis I) or by disuse (hypothesis II). Hypothesis I is based on the idea that osteoclasts are recruited and activated where microcracks occur in the bone¹⁵, possibly as an effect of signals from pre-apoptotic osteocytes¹⁶. The dynamic forces of daily living are known to produce microcracks²⁴. As the trabecular bone structure is a mechanically optimal one, implying that all material is frequently stressed in daily life²⁵, and the strength of the mineralized tissue is non-homogeneously distributed²⁶, it is likely that microcracks or diffused damage can occur anywhere, at any time, for a normally functioning individual²⁷; in other words, the distribution of microcracks is spatially random. Hence, for hypothesis I the probability of resorption, p , would be equal for all surface sites, and independent of mechanical strain²⁸; whereas for hypothesis II the local osteoclast-activation probability, p , would be high in disused areas and low in areas of high strain. Osteoclasts are activated by cytokines, which are produced by osteoblasts, and hence probably also by lining cells at the bone surface²⁹. This activation process could originate from a signal of mechanical disuse in the bone matrix, sent by osteocytes^{14,16}. In reality it is likely that both factors, disuse and microcracks, act concurrently in attracting and activating osteoclasts. They could work through the same signalling pathway, if we assume that the osteocytic network normally suppresses osteoclast activation by transporting signals to the surface through mechanical

loading. Disuse would then hamper the suppression by lack of loading, while microcracks would produce a similar effect through disconnection of the canaliculi¹⁴.

The local change of relative bone density m (1.0 for fully mineralized tissue) can then be expressed as

$$\frac{dm}{dt} = \tau\{P(x, t) - k_{tr}\} - r_{oc} \quad \text{for } P(x, t) > k_{tr}$$

$$\frac{dm}{dt} = -r_{oc} \quad \text{for } P(x, t) \leq k_{tr}$$

where r_{oc} is the relative amount of mineral resorbed by osteoclasts per day in the volume considered; $P(x, t)$ (in $\text{mol mm}^{-2}\text{day}^{-1}$) is the osteoblast recruitment stimulus at surface location x as a function of time t ; k_{tr} is its threshold level; and τ is a proportionality constant. The osteoblast recruitment stimulus is derived from

$$P(x, t) = \sum_{i=1}^n f_i(x) \mu_i R_{ii}(t)$$

where $R_{ii}(t)$ (in $\text{J m}^{-3}\text{s}^{-1}$) is the SED rate in the location of osteocyte i ; μ_i is the mechanosensitivity of osteocyte i ; f_i is an exponential decay function⁶ describing the attenuation of the bone formation stimulus between osteocyte i and surface location x ; and n is the number of osteocytes in the neighbourhood of the surface location considered.

The equations were transformed to numerical algorithms, and combined with an FEA code. The computer simulation was conducted as an iterative process, during which the relative density m was regulated per element between 0.01 (no bone) and 1.0 (fully mineralized bone). Per iteration, a bone-surface patch had a p per cent chance of becoming subject to resorption according to

hypothesis I: $p(x, t) = 10\%$ (spatially random)

hypothesis II: $p(x, t) = c[a - P(x, t)]$, if $P < a$ (strain dependent)

$$p(x, t) = 0, \quad \text{if } P \geq a$$

where c and a are constants. The probabilities were effected in the model with a Monte Carlo generator. When activated, each osteoclast was assumed to remove a fixed amount of mineral. Per iteration, the FEA code provided the SED rate values in the osteocyte locations. We applied this to a two-dimensional model of a 2×2 mm piece of bone, loaded at the edges by a sinusoidal stress, cycling between 0 and 2 MPa, at 1 Hz.

When assuming osteoclast-activation probability to be independent of strain (hypothesis I), the simulation produced a homeostatic trabecular architecture, when started from an arbitrary initial

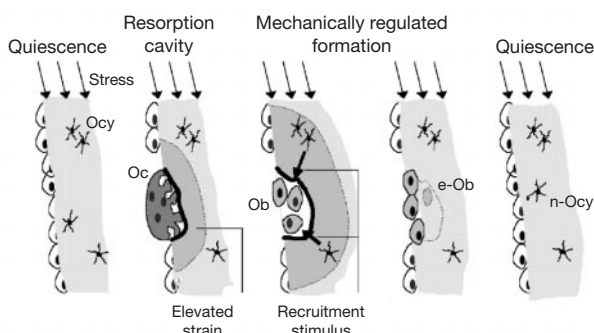


Figure 2 The trabecular surface is covered by lining cells (LC) and the osteocytes (Ocy) are inside the mineralized tissue. When an osteoclast (Oc) is recruited to resorb bone, a cavity is made which weakens the trabecula and causes a local elevation of strain. After the osteoclast has gone, osteoblasts (Ob) are thought to be recruited by osteocytes to form bone. During the bone formation process, some of the osteoblasts are entrapped in the bone matrix (e-Ob), where they differentiate to new osteocytes (n-Ocy). After repair, the remaining osteoblasts become lining cells, covering the new bone surface.

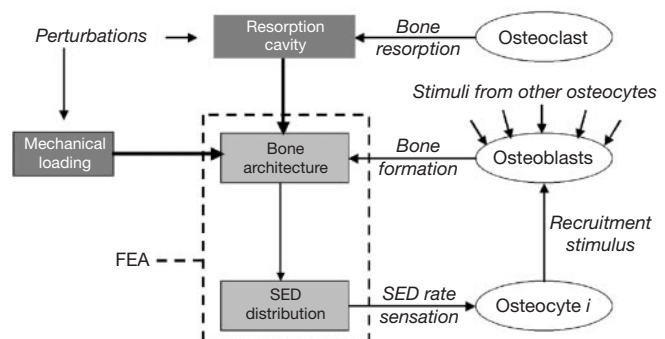


Figure 3 The proposed regulatory process (see text). Both enhanced external load intensity (amplitude and frequency) and resorption cavities provoke bone formation. Reductions in loading reduce bone mainly through unbalanced resorption.

configuration (Fig. 4a). In homeostasis, the process of remodelling by osteoclast resorption and osteoblast formation continues renewing bone, without affecting mass or architecture at large. When started from another arbitrary configuration (Fig. 4b), a similar architecture was eventually obtained, which confirmed the stability of the process, and established strain-related cell stimuli as effective control variables for architectural optimization. When the external load applied to the homeostatic architecture was rotated by 30°, the orientation of the trabeculae gradually reorientated as well, to align again with the external load (Fig. 4c). This is a distinct characteristic of trabecular architecture¹, in accordance with mechanical optimality. A 20% reduction in external loading reduced trabecular thickness to a loss of bone mass by 15.8%, which is of the order of that in individuals subjected to disuse¹⁰. Conversely, a 20% increase in loads produced a 17.5% increased bone mass, which is of the order of effects found from high-impact gymnastics⁹. When loading intensity was set back to original values, the original homeostatic configuration gradually re-emerged.

The only effect of strain-regulated resorption (hypothesis II), as opposed to the spatially random resorption described above, was in the kinetics of the process. The homeostatic architecture was maintained (Fig. 4d) but developed much faster in the simulation. This makes sense, as the osteoclasts help the osteoblasts to shape the form in the latter case, whereas they do not in the former case.

The parameters in the model can be linked to metabolic factors. The total mass in the homeostatic architecture, at any time, is the result of a balance between resorption and formation. If the amount of bone resorbed per volume per day (r_{oc}) remains constant, homeostatic bone mass is determined by the (typical) external loading magnitude and frequency, the osteocyte mechanosensitivity (μ_i), the osteocyte numbers (n), the stimulus distance attenuation function⁶ (f_i) and the osteoblast recruitment threshold (k_{tr} , the "set point"¹¹). On the other side of the balance, homeostatic bone mass is determined only by the total amount of bone resorbed per time, in the volume considered, which depends on osteoclast-activation frequency and amount of bone resorbed per osteoclast. When the osteoclast-activation frequency (r_{oc}) was increased by 50%, as can occur after menopause in women, trabecular thickness reduced, until a new balance was found at a 26% lower bone mass. The time curve obtained was similar to the one reported from epidemiological studies of bone mineral density development in women after menopause.

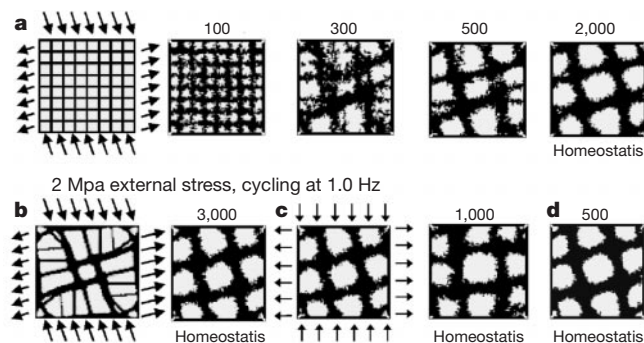


Figure 4 Results of the computer simulation. **a**, Iterative steps out of the simulation process, starting from a regular grid (spatially random resorption (hypothesis I): bone is black, marrow is white; iteration numbers are indicated). A homeostatic configuration is obtained, in which bone remodelling continues without changing the architecture at large. Trabecular orientation lines up with the external load. **b**, Starting from another initial configuration, the eventual homeostatic architecture is similar. **c**, After rotating the external loads, the architecture reorientates until adapted to the new loading directions. **d**, Homeostatic architecture after strain-controlled resorption (hypothesis II, $a = 1.6$, $c = 12.5$).

We conclude that mechanical feed-back, the existence of which is undeniable, can be a potent and stable regulator of the complex biochemical metabolic machinery towards lasting optimality of form. The regulatory paradigm that forms the basis of our model is relatively simple and amazingly stable. This again confirms the principles governing the emergence of complex consistent forms in nature under the influence of environmental stimuli, speculated upon long ago², but only now understood and made operational through the application of large-scale computer simulation⁷—the "third method of science"³⁰. □

Received 21 December 1999; accepted 9 March 2000.

- Wolff, J. *Das Gesetz der Transformation der Knochen* (Hirchwald, Berlin, 1892); translated as *The Law of Bone Remodeling* (trans. Maquet, P. & Furlong, R.) (Springer, Berlin, 1986).
- Thompson, D. W. *On Growth and Form* (Cambridge Univ. Press Cambridge, 1919).
- Roesler, H. The history of some fundamental concepts in bone biomechanics. *J. Biomechanics* **20**, 1025–1034 (1987).
- Cowin, S. C. Wolff's law of trabecular bone architecture at remodeling equilibrium. *J. Biomech. Engr.* **108**, 83–88 (1986).
- Carter, D. R., Fyhrie D. P. & Whalen R. T. Trabecular bone density and loading history: Regulation of connective tissue biology by mechanical energy. *J. Biomechanics* **20**, 785–794 (1987).
- Mullender, M. G. & Huiskes, R. A proposal for the regulatory mechanism of Wolff's law. *J. Orthop. Res.* **13**, 503–512 (1995).
- Stewart, I. *Life's Other Secret* (Allen Lane/Penguin, London, 1998).
- Lanyon, L. E. Using functional loading to influence bone mass and architecture: objectives, mechanisms and relationship with estrogen of the mechanically adaptive process in bone. *Bone* **18**, 375–435 (1996).
- Courteix, D. *et al.* Effects of physical training on bone mineral density in prepubertal girls: a comparative study between impact-loading and non-impact-loading sports. *Osteopor. Int.* **8**, 152–158 (1998).
- Zerwekh, J. E., Ruml, L. A., Gottschalk, F. & Pak, C. Y. C. The effects of twelve weeks of bed rest on bone histology, biochemical markers of bone turnover, and calcium homeostasis in eleven normal subjects. *J. Bone Min. Res.* **13**, 1594–1601 (1998).
- Frost, H. M. Vital biomechanics. Proposed general concepts for skeletal adaptation to mechanical usage. *Calcif. Tissue Int.* **45**, 145–156 (1987).
- Rodan, G. A. Mechanical loading, estrogen deficiency, and the coupling of bone formation to bone resorption. *J. Bone Miner. Res.* **6**, 527–530 (1991).
- Chambers, T. J. The direct and indirect effects of estrogen on bone formation. *Adv. Organ Biol.* **5B**, 627–638 (1998).
- Burger, E. H. & Klein-Nulend, J. Mechanotransduction in bone—role of the lacuno-canalicular network. *FASEB J.* **13**, S101–S112 (1999).
- Skerry, T. M., Bitensky, L., Chayen, J. & Lanyon, L. E. Early strain-related changes in enzyme activity in osteocytes following bone loading in vivo. *J. Bone Miner. Res.* **4**, 783–788 (1989).
- Noble, B. S., Stevens, H., Reeve, J. & Loveridge, N. Identification of apoptotic changes in osteocytes in normal and pathological human bone. *Bone* **20**, 273–282 (1997).
- Smit, T. S. & Burger, E. H. Is BMU-coupling a strain-regulated phenomenon? *J. Bone Miner. Res.* **15**, 301–307 (2000).
- Van Rietbergen, B., Weinans, H., Huiskes, R. & Odgaard, A. A new method to determine trabecular bone elastic properties and loading using micro-mechanical finite-elements methods. *J. Biomechanics* **28**, 69–81 (1995).
- Mosley, J. R. & Lanyon, L. E. Strain rate as a controlling influence on adaptive modeling in response to dynamic loading of the ulna in growing male rats. *Bone* **23**, 313–318 (1998).
- Turner, C. H., Owan, I. & Takano, Y. Mechanotransduction in bone: role of strain rate. *Am. J. Physiol.* **269**, E438–E442 (1995).
- Rubin, C. T. & Lanyon, L. E. Osteoregulatory nature of mechanical stimuli: function as a determinant for adaptive bone remodeling. *J. Orthop. Res.* **5**, 300–310 (1987).
- Klein-Nulend, J. *et al.* Sensitivity of osteocytes to biomechanical stress in vitro. *FASEB J.* **9**, 441–445 (1995).
- Cowin, S. C., Moss-Salentijn, L. & Moss, M. L. Candidates for the mechanosensory system in bone. *J. Biomech. Eng.* **113**, 191–197 (1991).
- Burr, D. B., Forwood, M., Fyhrie, D. P., Martin, R. B. & Turner, C. H. Bone microdamage and skeletal fragility in osteoporosis and stress fractures. *J. Bone Miner. Res.* **12**, 6–15 (1997).
- Van Rietbergen, B., Müller, R., Ulrich, D., Rüeggsegger, P. & Huiskes, R. Tissue stress and strain in trabeculae of a canine proximal femur can be quantified from computer reconstructions. *J. Biomechanics* **32**, 443–451 (1999).
- Choi, K. & Goldstein, S. A. A comparison of the fatigue behavior of human trabecular and cortical bone tissue. *J. Biomechanics* **25**, 1371–1381 (1992).
- Vashishth, D. *et al.* In vivo diffuse damage in human vertebral trabecular bone. *Bone* **26**, 147–152 (2000).
- Pazzaglia, U. E., Andriani, L. & Di Nucci, A. The effects of mechanical forces on bones and joints. *J. Bone Joint Surg.* **79B**, 1025–1030 (1997).
- Chambers, T. J. & Fuller, K. Bone cells predispose bone surfaces to resorption by exposure of mineral to osteoclastic contact. *J. Cell Sci.* **6**, 155–165 (1985).
- Kelly, J. The third culture. *Science* **279**, 992–993 (1998).

Acknowledgements

This work was sponsored in part by the Medical Sciences section of the Netherlands Organization for Research (NWO).

Correspondence and requests for materials should be addressed to R.H. (e-mail: rik@orthp.azn.nl).

Biological sensing of small field differences by magnetically sensitive chemical reactions

James C. Weaver*, Timothy E. Vaughan* & R. Dean Astumian†

* Harvard–M.I.T. Division of Health Sciences and Technology, Massachusetts Institute of Technology, Rm. 16-319, Cambridge, Massachusetts 02139, USA

† Departments of Surgery, and of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA

There is evidence that animals can detect small changes in the Earth's magnetic field by two distinct mechanisms, one using the mineral magnetite as the primary sensor and one using magnetically sensitive chemical reactions^{1–14}. Magnetite responds by physically twisting^{2,15}, or even reorienting the whole organism in the case of some bacteria¹⁶, but the magnetic dipoles of individual molecules are too small to respond in the same way. Here we assess whether reactions whose rates are affected by the orientation of reactants in magnetic fields could form the basis of a biological compass. We use a general model, incorporating biological components and design criteria, to calculate realistic constraints for such a compass. This model compares a chemical signal produced owing to magnetic field effects with stochastic noise and with changes due to physiological temperature variation¹⁷. Our analysis shows that a chemically based biological compass is feasible with its size, for any given detection limit, being dependent on the magnetic sensitivity of the rate constant of the chemical reaction.

The model (Fig. 1) is based on three biological components: (1) a magnetically sensitive chemical reaction with a ligand product molecule L, described by a magnetic field-dependent, first-order rate constant $k_M(B)$; (2) an effective local volume, V_{local} , with an excess of neural receptors for L; and (3) a passive permeability that removes L to a downstream, degradative sink. This permeability can

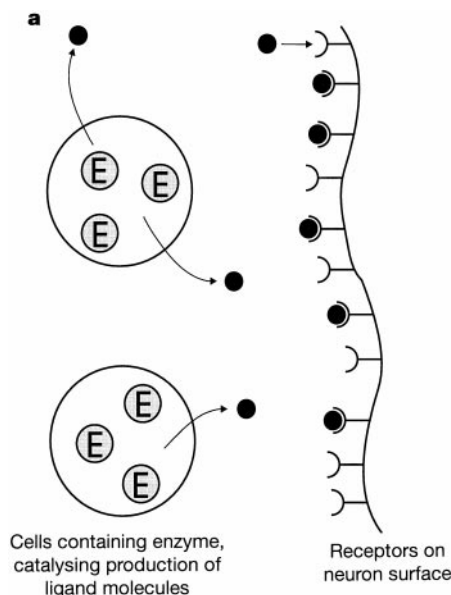


Figure 1 Key features of the model for biological detection of small magnetic field differences, showing what could be accomplished using biological components. The model also identifies essential features and constraints for any biological response based on radical-pair reactions. **a**, The detector is based on three components: (1) a magnetically sensitive radical-pair reaction that produces ligand L (filled circles); (2) an effective local volume with an excess of receptors for L; and (3) a distributed, passive permeability that removes L (not shown). Cells (large circles) contain enzyme (E) that

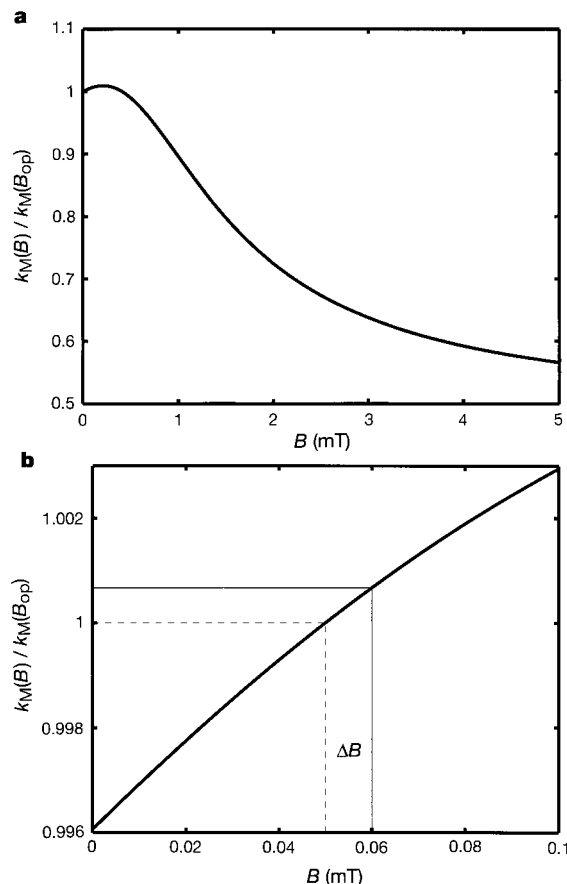
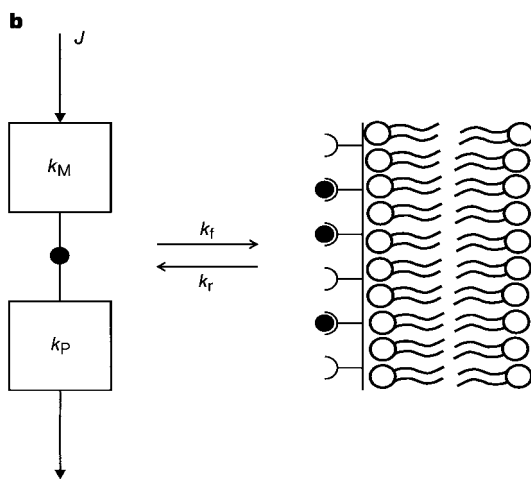


Figure 2 Fractional change in the reaction rate of the radical-pair reaction versus the magnitude of the magnetic field. **a**, The dependence for large field changes. **b**, The smaller changes expected from weak field anomalies. The dashed line indicates the 'operating point', the Earth's field; the solid line indicates the shift (due to the anomaly) from the operating point to the actual field. For clarity, the case of $\Delta B = 10^{-5}$ T (an extremely large anomaly, $\sim 20\%$ of the Earth's field) is shown.



catalyses the magnetic field-sensitive radical-pair reaction; the product ligand leaves the cell and establishes a local equilibrium between free and bound ligand (equation (2)). **b**, Diagram showing how the model parameters relate to components of the model. J is the total biochemical flux through the sensor. All other components are represented by their respective first-order rate constants: the magnetically sensitive radical pair reaction (k_M), downstream diffusive permeation (k_P), forward binding (k_f) and reverse binding (k_r).

Box 1

Molecular change signal-to-noise ratio

A detection threshold is estimated by comparing a signal with system noise. Here the signal S is the shift in the average number of complexes:

$$S \equiv \bar{C} - \bar{C}_{\text{op}} \approx g_1 \Delta B \frac{q_{\text{op}}}{(1 + q_{\text{op}})^2} R_T \quad (\text{A.1})$$

The noise, N , is the bound receptor fluctuation about the operating point:

$$N \equiv \delta C_{\text{op}} \approx \sqrt{\frac{q_{\text{op}}}{(1 + q_{\text{op}})^2} R_T} \quad (\text{A.2})$$

The signal-to-noise ratio is therefore

$$\frac{S}{N} = g_1 \Delta B \sqrt{\frac{q_{\text{op}}}{(1 + q_{\text{op}})^2} R_T} \quad (\text{A.3})$$

A common detection criterion is $S/N \geq 1$. The required number of receptors is then

$$R_T = \frac{(1 + q_{\text{op}})^2}{q_{\text{op}}} [g_1 \Delta B]^{-2} \quad (\text{A.4})$$

which scales the sensory size and is minimized for $q_{\text{op}} = 1$, when half of the receptors are occupied on average. We assume the evolutionary pressure optimizes the system. For the illustrative case of the Schulten–Schulten radical-pair theory, we find $R_T = 4(g_1 \Delta B)^{-2} \approx 4 \times 10^{10}$ receptors are required to detect an anomaly of $\Delta B = 10^{-7}$ T; if the anomaly is $\Delta B = 10^{-6}$ T (2% of the Earth's field), then only 4×10^8 receptors are needed.

be modelled as another first-order reaction, with a rate constant k_p that is independent of B . A metabolically driven biochemical flux J passes through components (1) and (3) such that c_L , the concentration of L in V_{local} , depends on the magnetic field magnitude. Because the average number of ligand–receptor complexes (C) depends on c_L through a rapid equilibrium between free and bound ligands, changes in the local magnetic field can be coupled chemically to the nervous system, an important part of a complete sensory system. Here we concentrate on the limiting transduction step: a field-induced molecular change.

For animal navigation based on detection of small direct current (DC) magnetic field anomalies, we describe the magnetic field by $B(t) = B_{\text{earth}} + \Delta B$, where $B_{\text{earth}} \approx 5 \times 10^{-5}$ T is the ‘operating point’ of the detector, and $\Delta B \ll B_{\text{earth}}$ represents a field anomaly. Widespread human use of electricity, predominantly in the form of alternating current at 50–60 Hz, has created time-dependent (oscillating) localized magnetic fields which can be regarded as anthropogenic magnetic anomalies. Sensory systems have, however, evolved to respond to steady DC anomalies^{10,12}. Some animals detect time-independent anomalies as small as 10^{-7} T (ref. 18).

Field alteration of the biochemical flux is described by $J = c_S k_M(B)$, where c_S is the substrate concentration of the chemical reaction. The magnetically sensitive chemical reaction is described by expanding $k_M(B)$ to second order about the operating point, B_{earth} :

$$k_M(B) \approx k_{M,\text{op}} (1 + g_1 \Delta B + g_2 [\Delta B]^2) \quad (\text{1})$$

where ‘op’ indicates the operating point value. A similar approach can be taken for directional sensing based on oriented molecules such as photoreceptors^{3,6,14} in this case expanding $K_M(B|\theta)$ in terms of $\Delta\theta$. For an illustrative radical-pair reaction theory¹⁹, the parameters $g_1 \approx 100 \text{ T}^{-1}$ and $g_2 \approx -3 \times 10^5 \text{ T}^{-2}$ were chosen to maximize the chemical change caused by the magnetic field, on the grounds that evolutionary pressure leads to optimal sensory performance. Figure 2a shows the field dependence of k_M for large fields, from 0 – 5×10^{-3} T (0 – 50 G), and Fig. 2b shows the dependence for a small range around a typical local value of the Earth's field, from 0 – 10^{-4} T (0 – 1 G).

Box 2

Minimum detector volume

The density of neural receptors can be as high as $d_{\text{rec}} \approx 10^4 \mu\text{m}^{-2}$ (ref. 22), corresponding to an area per receptor of $A_{\text{rec}} \approx 10^{-16} \text{ m}^2$. The total area occupied by all the receptors is

$$A_{\text{tot}} = R_T A_{\text{rec}} \approx 4 \times 10^{-6} \text{ m}^2 \quad (\text{B.1})$$

to detect an anomaly of $\Delta B = 10^{-7}$ T. Representing a nerve as a cylinder with average radius $r_{\text{nerve}} \approx 1 \mu\text{m}$, the requisite total nerve length is

$$l_{\text{tot}} = \frac{A_{\text{tot}}}{2\pi r_{\text{nerve}}} \approx 0.6 \text{ m} \quad (\text{B.2})$$

From considerations of diffusion time, which for a sensory system is about 1 s, the ligands can occupy a volume that is located up to $r_{\text{max}} \approx 50 \mu\text{m}$ from the receptors. The required detector volume is therefore

$$V_{\text{tot}} = \pi r_{\text{max}}^2 l_{\text{tot}} \quad (\text{B.3})$$

Using the Schulten–Schulten theory as an example, $V_{\text{tot}} \approx 5 \times 10^{-9} \text{ m}^3$, which is less than about $(2 \text{ mm})^3$. In this case, the required volume is less than about $(0.4 \text{ mm})^3$ if the anomaly is $\Delta B = 10^{-6}$ T (2% of the Earth's field).

The operating point rate constant is $k_{M,\text{op}} = k_M(B_{\text{earth}})$. For $\Delta B = 0$, the flux is $J_{\text{op}} = c_S k_{M,\text{op}}$. The flux shifts slightly from J_{op} in response to a small ΔB (Fig. 2b). For small DC anomalies, the linear term dominates, but for 50–60 Hz fields, the molecular change proportional to ΔB averages to zero. The contribution proportional to $(\Delta B[t])^2$ describes the rectified response of the detector to 50–60 Hz fields, and is the source of any response to these man-made anomalies. However, most animal navigation takes place far from these sources, and the contribution of these background fields is therefore negligible.

The field dependence of the ligand concentration is

$$c_L = \frac{k_M(B)}{k_p} c_S \quad (\text{2})$$

A magnetic field anomaly changes the ligand concentration, and this in turn influences the number of ligand–receptor complexes. We assume that there is a rapid, local equilibrium between free and bound ligands, governed by the reactions



with equilibrium dissociation constant $K_D = k_r/k_f$. Table 2-1 in ref. 20 gives binding parameters of several ligand–receptor systems, including one with a characteristic response time of a few seconds, sufficiently rapid for navigational sensing.

Fundamental noise is the stochastic fluctuation of the number of complexes about an equilibrium average, $\bar{C}$, as described by the binomial distribution^{20,21}

$$\bar{C} = \frac{1}{1 + q} R_T \quad (\text{4})$$

Here R_T is the total number of receptors, and $q \equiv K_D/c_L$ is defined for convenience. Fundamental fluctuations in the number of complexes about $\bar{C}$ are characterized by the variance of the distribution, and are typically of order²⁰

$$\delta C = \sqrt{\frac{q}{(1 + q)^2} R_T} \quad (\text{5})$$

A detection threshold can be estimated using a signal-to-noise ratio criterion, $(S/N) \approx 1$, which leads to the condition for the number of receptors needed to overcome stochastic fluctuations (Box 1). This number is compatible with a small, multicellular system (Box 2).

Box 3

Protection against temperature variations

A significant restriction on a biological detector is that the molecular change due to the field must be greater than the molecular change due to temperature variations, ΔT . Temperature variations are inescapable, and typically modify biochemical reaction rates and transport properties according to $\Delta I \approx \alpha \Delta T$, a typical large value ($\alpha \approx 3\% \text{ } ^\circ\text{C}^{-1}$) for individual biochemical processes¹⁷. For the sensory detector system, the temperature coefficient is

$$\alpha_{\text{det}} = \frac{1}{C} \frac{\partial C}{\partial T} \quad (\text{C.1})$$

Using equations (2) and (4) and the definition of q , the temperature coefficient of the detector at the operating point is

$$\alpha_{\text{det}} = \frac{1}{2}(\alpha_P - [\alpha_M + \alpha_K]) \quad (\text{C.2})$$

where α_P , α_M and α_K are the temperature coefficients of the corresponding components of the model. To determine whether magnetic field effects are detectable in the presence of temperature variations, we compare the first-order changes due to ΔB and ΔT . Using equations (C.1) and (A.1), we find that in the presence of both magnetic field anomalies and temperature variations,

$$\bar{C} \approx \bar{C}_{\text{op}}(1 + \frac{1}{2}g_1\Delta B)(1 + \alpha_{\text{det}}\Delta T) \quad (\text{C.3})$$

On a time scale of seconds to minutes, *in vivo* variations for both cold- and warm-blooded animals often exceed $\Delta T \geq 0.1 \text{ } ^\circ\text{C}$. This leads to an estimate that a temperature coefficient of about

$$\alpha_{\text{det}} \leq 5 \times 10^{-5} \text{ } ^\circ\text{C} \quad (\text{C.4})$$

is required for temperature variations not to overwhelm magnetic field sensitivity to DC anomalies of the order of 10^{-7} T . This coefficient is small, but evolutionary pressure may be able to achieve it by matching α_P to approach $(\alpha_M + \alpha_K)$.

Significant temperature variations¹⁷ can be overcome (Box 3). We conclude that a magnetic field sense based on magnetically sensitive chemical reactions can overcome both fundamental fluctuations in ligand–receptor complexes, and competing molecular changes due to inescapable temperature variations. A small, multicellular biological detection system can provide both of these capabilities, and might also be able to improve sensing performance by neural processing, which we have not invoked here. This analysis

demonstrates quantitatively that an extremely sensitive, chemically based magnetic sense should be possible. □

Received 16 February; accepted 26 April 2000.

1. Leask, M. J. M. A physicochemical mechanism for magnetic field detection by migratory birds and homing pigeons. *Nature* **267**, 144–145 (1977).
2. Kirschvink, J. L. & Gould, J. L. Biogenic magnetite as a basis for magnetic field detection in animals. *Biosystems* **13**, 181–201 (1981).
3. Schulten, K. Magnetic field effects in chemistry and biology. *Festkörperprobleme* **22**, 61–83 (1982).
4. Beason, R. C. & Nichols, J. E. Magnetic orientation and magnetically sensitive material in a transequatorial migratory bird. *Nature* **309**, 151–153 (1984).
5. Phillips, J. B. Two magnetoreception pathways in a migratory salamander. *Science* **233**, 765–767 (1986).
6. Schulten, K. & Windemuth, A. in *Biophysical Effects of Steady Magnetic Fields* (eds Maret, G., Boccarda, N. & Kiepenheuer, J.) 99–106 (Springer, Berlin, 1986).
7. Phillips, J. B. & Borland, S. C. Behavioural evidence for use of a light-dependent magnetoreception mechanism by a vertebrate. *Nature* **359**, 142–144 (1992).
8. Able, K. P. & Able, M. A. Interactions in the flexible orientation system of a migratory bird. *Nature* **375**, 320–322 (1995).
9. Beason, R. C., Dussourd, N. & Deutschlander, M. E. Behavioural evidence for the use of magnetic material in magnetoreception by a migratory bird. *J. Exp. Biol.* **198**, 141–146 (1995).
10. Wiltschko, W. & Wiltschko, R. *Magnetic Orientation in Animals* (Springer, Berlin, 1995).
11. Walker, M. M. *et al.* Structure and function of the vertebrate magnetic sense. *Nature* **390**, 371–376 (1997).
12. Deutschlander, M. E., Phillips, J. B. & Borland, S. C. The case for light-dependent magnetic orientation in animals. *J. Exp. Biol.* **202**, 891–908 (1999).
13. Deutschlander, M. E., Borland, S. C. & Phillips, J. B. Extraocular magnetic compass in newts. *Nature* **400**, 324–325 (1999).
14. Ritz, T., Adem, S. & Schulten, K. A model for photoreceptor-based magnetoreception in birds. *Biophys. J.* **78**, 707–718 (2000).
15. Yorke, E. D. A possible magnetic transducer in birds. *J. Theor. Biol.* **77**, 101–105 (1979).
16. Blakemore, R. P., Maratea, D. & Wolfe, R. S. Isolation and pure culture of a freshwater magnetic spirillum in chemically defined medium. *J. Bacteriol.* **140**, 720–729 (1979).
17. Weaver, J. C., Vaughan, T. E. & Martin, G. T. Biological effects due to weak electric and magnetic fields: the temperature variation threshold. *Biophys. J.* **76**, 3026–3030 (1999).
18. Lohmann, K. J. & Lohmann, C. M. F. Detection of magnetic field intensity by sea turtles. *Nature* **380**, 59–61 (1996).
19. Schulten, Z. & Schulten, K. The generation, diffusion, spin motion and recombination of radical pairs in solution in the nanosecond time domain. *J. Chem. Phys.* **66**, 4616–4634 (1977).
20. Lauffenburger, D. A. & Linderman, J. L. *Receptors: Models for Binding, Trafficking and Signaling* (Oxford Univ. Press, New York, 1993).
21. van Kampen, N. G. *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
22. Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principles of Neural Science* 3rd edn (Appleton and Lange, Connecticut, 1991).

Acknowledgements

We thank J. C. Squire, A. Sastre, J. B. Phillips, G. T. Martin, S. K. Burns, R. K. Adair and E. Adair for discussions. Supported by the Massachusetts Institute of Technology Electric Utilities Program Consortium and by a computer equipment grant from Stadwerke Düsseldorf, Düsseldorf, Germany.

Correspondence and requests for materials should be sent to J.W. (e-mail: jim@hstbme.mit.edu).

More things to do with DNA

Blow air at it, heat it, extract it, repair it, spin it, amplify it, express it...

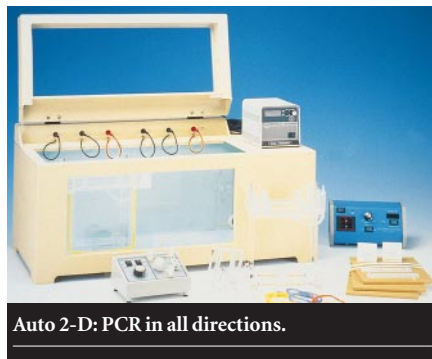
Auto 2-D TDGS

C.B.S. Scientific www.cbsscientific.com

Combining PCR and 2D electrophoresis

This automatic two-dimensional gene scanning system (TDGS) combines multiplex PCR with the high resolution of two-dimensional denaturing gradient gel electrophoresis. TDGS is designed for screening large numbers of individual DNA sample for mutations and polymorphisms in very large or multiple genes. Auto 2-D TDGS consecutively resolves DNA fragment mixtures on the basis of both size (1-D) and base pair sequence (2-D) within the same gel without manual interference.

Reader Service No. 100



Auto 2-D: PCR in all directions.

ods, chemistry hits, application-specific support and after-sales service.

Reader Service No. 102

DNA-free

Ambion www.ambion.com

Take the genomic DNA out of RNA

The latest issue of Ambion's *TechNotes* publication discusses a novel way of removing contaminating genomic DNA from RNA samples, making them suitable for RT-PCR. The 'DNA-free' DNase Treatment and Removal system includes a DNase Removal Reagent which, after digestion, quickly eliminates DNase without phenol extractions or heat inactivation procedures which can cause RNA loss or degradation. RNA isolation methods specifically for RT-PCR are also featured.

Reader Service No. 101

Biomek FX

Beckman Coulter, Inc.

www.beckmancoulter.com

Automation of DNA preparation with advanced liquid handling technology

Capabilities claimed for the Biomek FX Nucleic Acid Preparation system include speed, deck size and dual pipetting modes, allowing the automation of many DNA preparation functions, some not previously automated. The Biomek FX package incorporates hardware, software, prevalidated meth-



Acid test: Biomek FX automates the unautomated.

Anti-APE

Trevigen Inc. www.trevigen.com

A DNA repair and apoptosis related antibody

Trevigen have introduced a new affinity purified polyclonal antibody against human AP endonuclease (APE), also known as Ref-1. APE is a multifunctional enzyme involved in DNA repair and in facilitating the redox state of DNA binding proteins. Transcription factors that the enzyme acts on include FOS, JUN, NF-KB, and p53. APE expression is also associated with apoptosis induction. The antibody may be used for both immunocytochemistry and Western blotting applications.

Reader Service No. 103

GenPrep and GenSpin

Cambridge Molecular www.cmtech.co.uk

Genomic purification kits

GenPrep and GenSpin were developed for the rapid purification of single-stranded genomic DNA for PCR amplification. Both kits use a filter paper material designed to simplify sample collection and archiving, as well as achieving rapid purification of soluble genomic DNA from small samples. GenPrep uses a free-floating filter disk in a microtube to give hands-off purification in a single tube. Using up to 7.5 µl blood or 100 µl saliva, a simple heat elution step yields up to 330 ng soluble DNA. Higher quantities of DNA can be achieved with the GenSpin Kit, based on a spin cartridge technique and a standard bench-top centrifuge.

Reader Service No. 104

Anti-thymine dimer antibody

Affitech AS www.affitech.com

Monitor DNA damage and repair

H3 is an antibody with high affinity and specificity for thymine dimers, the dominant form

of chemical modification in DNA exposed to UV light. The presence of thymine dimers in DNA can be used to monitor the effect of UV light on cells, both *in vivo* and *in vitro*. H3 antibody can also be used to detect subpicogram amounts of UV-irradiated DNA directly in a non-radioactive detection system (*BioTechniques*, 23, 892–896; 1997). Following a short exposure to UV-light, the antibody can also be used to improve the sensitivity of DNA detection on membranes, for example in Southern blots.

Reader Service No. 105

MagNA Pure LC

Roche Molecular Biochemicals

www.roche.com

Just walk away

The MagNA Pure LC automated nucleic acid isolation and PCR set-up can isolate DNA, RNA or mRNA from up to 32 samples in about one hour. The instrument's pipetting system can be used directly to fill LightCycler capillaries or 96-well plates with purified nucleic acids and PCR reagents. Samples are ready for real-time PCR in Roche's LightCycler or conventional thermal cyclers.

Reader Service No. 106

Purefect

Matrix Technologies matrixtechcorp.com

Disposables covered in a new brochure

The Purefect range of disposables is guaranteed free of contaminants including bio-burden, endotoxin (pyrogen), DNA, DNase, RNase, and cytotoxicity. Testing for these



Brief lives: book to bin, quite quickly.

new on the market

impurities is performed free of charge. The brochure outlines the various purity tests that Matrix use and shows representative results. Purefect disposables include deep well tubes, cap mats, cap strips, reservoirs, tips, microplates and blocks.

Reader Service No. 107

NucleoVision 760

NucleoTech Corporation

www.nucleotech.com

Visible and fluorescence imaging and analysis

A complete imaging workstation designed for advanced densitometry analysis, the NucleoVision 760 takes advantage of the growing source of information based on images including electrophoretic gels, autoradiographs, blots, and even 3-D objects. The system offers densitometry and molecular weight calibration, data reproducibility, and automation via an interface designed with the molecular biologist in mind. Applications include 1-D gel analysis, protein gels, Southern/ northern, PCR, QPCR, colony/plaque counting, and microscopy.

Reader Service No. 108

mRNA isolation

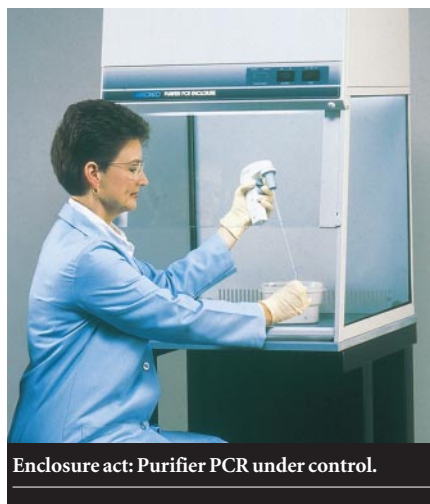
Miltenyi Biotec

www.miltenyibiotec.com

A new protocol for the isolation of mRNA directly from human whole blood

Using the MACS mRNA Isolation Kits in combination with a MACS separator, Miltenyi Biotec claim that mRNA can be isolated in just a few pipetting steps. In 15–30 minutes small or large scale isolation from 0.5 to 2.5 ml whole blood can be performed. The composition of the lysis buffer is optimized for efficient lysis of erythrocytes. Highly pure, functional mRNA is obtained, particularly useful for experiments such as RT-PCR or microarray screening.

Reader Service No. 109



Enclosure act: Purifier PCR under control.

Purifier PCR Enclosures

Labconco Corporation

www.labconco.com

Controlled environment for PCR

This bench-top system uses an integral blower and a 99.9% efficient HEPA filter to circulate filtered, Class 100 air down across the work area, providing a particulate-free work space to minimize the risk of contamination of the samples. Available in 2- and 3-foot widths, the Purifier PCR Enclosures feature a self-contained UV lamp with solid-state timer which provides a 5 minute exposure to deactivate DNA and RNA contaminants. The UV light then automatically switches off, in preparation of the next experiment.

Reader Service No. 110

Bactozol

Molecular Research Center, Inc.

www.mrcgene.com

A kit for bacterial DNA isolation

Central to the Bactozol kit for isolating DNA from a broad spectrum of Gram-negative and Gram-positive bacteria is Bactozyme. This is a concentrated solution (10×) containing an activated lysozyme stable at room tempera-

ture, avoiding daily enzyme preparation. After cell lysis in Bactozyme, DNAzol provides a rapid method for isolation of DNA from the bacterial lysate. The kit includes the reagents for 125 isolation procedures, each yielding up to 40 µg of DNA.

Reader Service No. 111

TruTemp

Robbins

www.robsci.com

Small-scale heating

The TruTemp DNA Microheating System consists of a heated chamber with lid, designed for fast heat up and consistent controlled temperatures. A dual heating system denatures DNA samples at 95°C from ambient in under 12 minutes. For hot start PCR protocols, reaction mixtures can be rapidly heated to the starting temperature, maintained at that temperature for the addition of the last reagent, and then moved to the thermocycler. Temperature control is $\pm 0.2^\circ\text{C}$.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers.



Open government: TruTemp controls to $\pm 0.2^\circ\text{C}$.

ADVERTISEMENTS

mtDNA EXTRACTOR KITS

NEW

- Easy & efficient isolation of mtDNA from whole blood, cell culture or tissue
- Results in ~90 Minutes
- No Hazardous Reagents
- Works even in the presence of anticoagulants

Wako

Wako USA Wako GmbH
877-714-1921 (49)02131 3110
www.wako-chem.co.jp

BACs • YACs • PACs cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse and rat libraries as well as a custom screening service for the libraries.

Research Genetics

U.S. or Canada 800-533-4363

U.K. 0-800-89-1393

FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

Look no further

for the best value in

Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!

SIGMA
GENOSYS

(+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 38

READER ENQUIRY NO. 33

Gender issue

The number of women scientists is still woefully small

p713

History lesson

The unsung women heroes of science provide inspiration

p715

Balancing act

New legislation in Britain is making slow progress towards equality

p716

In the minority

US ethnic minorities face an uphill struggle in scientific careers

p717

US science shocked by revelations of sexual discrimination

The scientific establishment in the United States has received a resounding wake-up call over sexual equality. Natasha Loder investigates.

The battle of the sexes is far from over. Two studies, one in Sweden¹, the other in the United States², have produced startling evidence that sexual equality in science remains a distant goal. But the results of these surveys are at last providing some impetus for change.

In Europe, researchers were shocked by the damning results of the Swedish study (see box overleaf). But any complacency in the United States was shaken with results that showed discrimination at one of its leading academic institutions.

Started in 1994, the study at the Massachusetts Institute of Technology (MIT) looked at the working conditions of senior female faculty members in the institute's School of Science. A summary report was published in the spring of last year³. The full report, issued to MIT's dean, documented a pattern of gender discrimination in hiring, promotion, awards, committee responsibility, the allocation of laboratory space and research money. It also found that, despite increasing numbers of women scientists, there had been no change in faculty ratios for 10–20 years (see chart overleaf). Commendably, MIT acknowledged the findings and introduced a number of swift remedies and sweeping administrative changes.

Ripples through the system

The study was chaired by molecular biologist Nancy Hopkins, a well-respected scientist working on the control of early vertebrate development⁴. For a year it totally overwhelmed her life — a year on, Hopkins says she can't be objective about the report's impact. "I travelled endlessly, spoke in about 100 places, talked to 30 or 40 newspapers and there was a TV 20-minute film. The dean was also buried. They say here that only Nobel prizes cause this much stir," she says. The news made a huge splash in the national and international media — and the waves were felt by administrators in academic institutions around the nation.

Within six months, Case Western Reserve University, Harvard Medical School,



Hopkins: her report of sexual discrimination against women scientists has catalysed change.

the University of Arizona, the California Institute of Technology (Caltech) and the University of California at Los Angeles (UCLA) all had plans to study gender equality as a result of the MIT report⁵. Hopkins says she has lost count of the number of enquiries she has had about the study.

The MIT report "has made a lot of places look at themselves", says Anneila Sargent, a professor of astronomy and director of the Owens Valley Radio Observatory at Caltech. Sargent is chairing a study into overall job satisfaction for both male and female faculty members.

The University of Arizona has set up a Millennium Project to look at the working conditions for all female professors, and at UCLA the MIT report has nudged a study of gender inequality into action⁶. One dean at an Ivy League university who found no grass roots interest for a review of gender equality reportedly set one up "just to be sure".

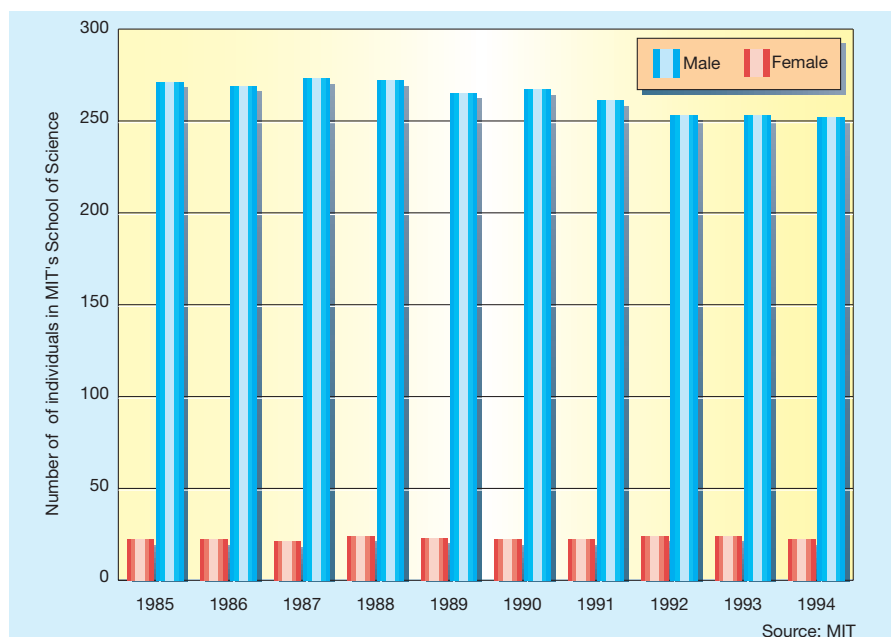
The effects of the MIT report have been "profound", says Helen Davies, a microbiologist at the University of Pennsylvania Medical Center. "It rates as one of the most important reports on academic women in

science in the past two decades because it concerns one of the most important research universities and resulted in an admission of discrimination," she says. Davies has served for two years as president of the Association for Women in Science and has, in the words of Hopkins, "been doing for 20 years what we've been doing for five".

The report has even made an impact on universities that thought they were safe. When Davies initially read the MIT report, she was astounded that MIT was so far behind — the University of Pennsylvania had conducted a similar investigation in 1970. "For a while their motto was 'We've been there, done that and bought the T-shirt'," Davies says of Pennsylvania. "Then we went out and did a preliminary study on three-year-old data. We did not look appreciably better than MIT, though we started 29 years ago. This was a shock."

The shocks continue to come. This year Davies notes an "historic" admission by the University of Rhode Island that its engineering school was hostile to women. And a study in the *New England Journal of Medicine* showed that women who teach in medical

RICK FRIEDMAN



The imbalance of men and women in the School of Science at Massachusetts Institute of Technology.

► schools are less likely to be promoted at every step along the path⁶. “Whatever studies are done, this is what you see,” says Davies.

There are worries that some institutions still have their heads in the sand. Some, such as Harvard⁷ and Stanford, have said that they do not have this problem. “Of course, women in these institutions tell us they do have a problem,” Davies says. “Everywhere I go, women tell me problems exist, and the issues are virtually identical.”

There is no way of knowing how many women face these problems, but Hopkins says she has found inequality in “every institution I have been to”. She adds: “In some fields where there are really few women, like engineering, women are driven away and they can’t even put their finger on what it is that is unfriendly.”

Hostile territory

Some fields are more hostile to women than others, agrees Judy Franz, executive officer of the American Physical Society (APS). Franz heads a committee that, when invited, visits physics departments around the country and assesses their environment. The programme, which started in 1990, has visited about 15 departments and will soon start evaluating federal labs. “Early on, we would hear a lot about individual indignities,” Franz says. For example, in one department, women graduate students would be asked to substitute for secretaries, while males were not.

Things have improved slowly, although the number of women physicists has not grown drastically. The APS visits can make a difference, Franz says. In one instance, a team found that all physics graduate students were housed in a separate building with an atmosphere she describes as “Animal House”, complete with pictures of scantily clad females on

the walls. After the APS team reported the situation, the students were moved back into the departments and the posters disappeared.

Not everyone agrees that discrimination against women scientists is prevalent. The Independent Women’s Forum says that without publishing the raw data from the full MIT report, the available summary serves as no more than a political manifesto.

Hopkins, who first mistook the complaint as a joke, explains that most of the data were only given on the basis of confidentiality. “The actual report has all types of individual studies and a ton of private information,” she says. “That has never been published and never will.”

Difficulty obtaining data is a common theme in studies of this kind. Access to salary

data is the first hurdle that the study at Caltech is struggling with. And the Swedish study was only possible after a long battle involving lawyers and the country’s freedom-of-information laws.

Hopkins also warns that the MIT report was not a statistical survey. “It was taking a lot more data than is ever gathered in making professional judgements,” she says. Normally, decisions on salary or how much space an employee gets are “made by one man who sits alone in a room”, she adds.

Effects on the ground

After 30 years in science, Caltech’s Anneila Sargent talks of a “climate change in the whole of academia”, and notes changes in the start-up packages women are getting. The threat of lawsuits might also be focusing the minds of university administrators.

At Pennsylvania in February, a women’s faculty and administration held a meeting for those interested in the MIT study. Instead of the usual 15 or 20 women, 150 turned up. The interest was “wonderful, absolutely tremendous” says Davies. Similar interest was reported at the University of Texas⁸.

Hopkins describes MIT’s transformation as “extraordinary”. She says the cohesiveness of the group of women that wrote MIT’s report was a crucial factor in its impact. But maybe the weight of evidence could simply no longer be ignored. ■

Natasha Loder is a news reporter at Nature.

1. Wennerås, C. & Wold, A. *Nature* **387**, 341–343 (1997).
2. Zernike, K. MIT women win fight against bias. *Boston Globe* 21 March (1999).
3. *The MIT Faculty Newsletter* special edition XI, 4 (1999). <http://web.mit.edu/fnl/women/women.html>
4. Nadis, S. *Nature* **398**, 361 (1999).
5. Wilson, R. An MIT professor’s suspicion of bias leads to a new movement for academic women. *Chronicle of Higher Education* 3 December (1999).
6. *New Eng. J. Med.* **342**, 399–405 (2000).
7. Lawler, A. *Science* **286**, 1272–1278 (1999).
8. Loder, N. *Nature* **402**, 337 (1999).
9. Loder, N. *Nature* **400**, 202 (1999).

Europe faces up to its responsibilities

Published in 1997, a study in Sweden showed that women applying for funding from the Swedish medical research council needed more than twice the qualifications of their male counterparts to receive support¹. This result surprised many, as Sweden is known for its equality of opportunity for women. Before this study, sceptics in Europe had treated complaints of discrimination against women scientists as mere anecdotes, easily brushed aside. Now they were forced to act.

The Massachusetts Institute of Technology’s report on discrimination against women was important, not as a catalyst for change, but as an added impetus to the moves in Europe driven by the Swedish study. Today, one of the outcomes can be seen in a report produced by the European Technology Assessment Network (ETAN), a science policy group established by the European Commission.

The ETAN report represents the first systematic attempt to gather statistics on European women in science². It concludes that under-representation of women and old-fashioned employment practices are undermining efforts to achieve excellence in European science. Currently, the report is playing an important role in driving forward political action, and activists have been relieved that Philippe Busquin, the new Commissioner for Research, is as keen as his predecessor Edith Cresson to monitor and improve the situation in member states.

N. L.

General outline of European action ♦ http://www.cordis.lu/improving/src/women_outline.htm

ETAN report ♦ http://www.cordis.lu/improving/src/hp_women.htm

New tenure-track programme in a German university ♦ http://www.fz-juelich.de/pr/jobs-extern/tenure_e.html

Making moves to redress the gender imbalance

The name Herschel is well known to astronomers. A description of the observations of nebulae and the cataloguing efforts of both Sir William Herschel (1738–1822) and his son Sir John (1792–1871) appears in *The World Treasury of Physics, Astronomy and Mathematics*. But curiously there is no mention of William's sister Caroline.

Caroline Herschel (1750–1848) is considered to be the first noted woman astronomer of the modern age. She made original observations, re-catalogued Flamsteed's star catalogue and was employed as assistant to the King's astronomer for £50 per year — which at the time was a very reasonable salary.

In 1828, the Royal Astronomical Society made her an honorary member — women were not allowed to be members. The Royal Irish Academy elected her a fellow, and the King of Prussia awarded her a gold medal for science. Yet Caroline is not mentioned in *The World Treasury*, a telling omission for a book published in 1991 in the United States. She is, after all, a woman who broke through the glass ceiling in spectacular fashion.

Caroline's invisibility today in the popular presentation of the historical record is, it seems, symptomatic of an ethos that still accords lower status to women scientists than it does to men. The evidence of the disparity in status in academia between men and women is overwhelming.

Mind the (salary) gap

An independent review in 1999 of pay and conditions in UK higher education — known as the Bett Report — found that in both full- and part-time posts, the average salary for women was markedly less than that for men. Also, women were more likely than men to be on fixed-term contracts rather than in permanent posts. The Bett Report's findings on salary echo those of a survey in 1998 by the UK Institute of Physics, which found that women had a median salary of £21,000 (US\$31,400) compared with £30,000 for men.

Just to drive the point home, the National Association of Teachers in Further and Higher Education (NATFHE) recently assembled a league table using the latest figures available (1997–1998) from the Higher Education Statistics Agency. The table showed women professors of anatomy and physiology receiving £8,000 per year less than their male counterparts. There were no women professors at all in civil engineering.

This status quo is now under attack. Since the 1993 UK government policy paper

"Realising our potential" identified women as the single most underused resource in society, there has been a lot of activity. Higher-education institutions, trade unions and successive Conservative and Labour governments have all sponsored initiatives aimed at attracting women to science, engineering and technology and then providing an environment that retains them.

Same as it ever was

There has not yet been time for such initiatives to show progress and, judging by the debate on the subject sponsored last November by the Parliamentary Office of Science and Technology (POST), there is still a long way to travel. Much of what participants said about the low numbers of women in science, the need for role models, mentors and less stereotyping of the scientists' image, could easily have been said 20 years ago.

"I was required to attend a part of the British Association's annual festival in Sheffield [in 1999]," says Janet Drew, professor of astrophysics at Imperial College, London. "The predominance of grey-haired gentlemen in jackets and ties depressed even me. What must schoolchildren attending such events think?"

After the POST debate had run for three weeks,

the Women in Higher Education Register, which is funded by the Committee of Vice-Chancellors and Principals, collated recommendations. These included encouraging institutions to adopt flexible working practices; developing mentoring and role-model schemes to encourage women to continue with science; gathering data on women scientists' career advancement; and establishing a broader range of criteria for promotion than the number of papers published.

This last recommendation is an implicit criticism of the UK's Research Assessment Exercise (RAE) — and the same point was made by Sir Gareth Roberts at Sheffield University in the May report of the Research Careers Initiative (see related story overleaf). The RAE evaluates the quality of a university and so determines the institution's access to funding. Several participants in the POST debate argued that assessments should be based not only on the number of publications, but also on team working, communication skills, teaching and contributions to interdisciplinary work.

Even before these recommendations appeared, a number of institutions were thinking along similar lines. It is just that currently, says Bob Ward of the Science Policy Unit of the Royal Society, the approach is piecemeal. One specific initiative is the Society's Dorothy Hodgkin Fellowship, which offers young

Caroline Herschel, the forgotten astronomer.



SCIENCE PHOTO LIBRARY

Table 1 Percentage of women chemists in the United Kingdom by grade

Grade	% Female staff		
	All institutions	Established universities	Former polytechnics and colleges
Professor	<1	<1	0
Senior lecturer	4	4	6
Lecturer	13	12	16
Researcher	22	22	32
Other staff	25	27	22
All	16	16	16

scientists a four-year position, along with help to allow them to raise a family and simultaneously pursue a career. Support includes freedom to switch between full- and part-time work, and help with child-care expenses while attending overseas conferences.

While the fellowships are open to men and women, the scheme is structured to be more attractive to women than men. Vice-chancellors are asked to make the scheme known in departments that traditionally have very few women. So far, says the Royal Society's Keith Wylde, the strategy has attracted many applicants, more than 80 per cent of whom are women. Currently, all but one of the 42 Fellows are women.

Chemical imbalance

As the Royal Society makes a modest start to the task of redressing the imbalance between men and women in higher education, the Royal Society of Chemistry (RSC) is working to understand the choices that male and female chemists make. It released findings of a survey earlier this year (see websites).

The study found that women are poorly represented in higher-education research in general but especially so in chemistry. The percentage of women academic chemists falls from 33 per cent at postgraduate level to less than 1 per cent at professorial level — in the UK sense of a professorship (see Table 1).

This leaching away of women from the first rungs of the academic career ladder is not unique to chemistry, and seems to be associated largely with women taking career breaks to have children, then finding that their publication rate has fallen. Women, too, still have the lion's share of child-care responsibilities.

For women working at the laboratory bench and needing to put in long hours, as in chemistry, child-care responsibility can make a life in academia very difficult. Given that academic research is also poorly paid, by the time a woman has paid for professional child care she is unlikely to have much in her pay packet.

The RSC's study notes that the prospects of reversing the effects of women leaving academic chemistry are poor. On current trends it would take until 2070 for women to reach parity with men.

The prospects for women in chemistry in the immediate future are particularly dim,

says the study, because there has been a squeeze on the number of middle-ranking posts in chemistry. In addition, a recent set of 46 new appointments to professorships were all awarded to men.

But women chemists do get jobs in chemistry. So why do they not stay in academia? The reasons are complex, but attitudes play a part. The RSC study found that men see chemistry as hard-edged and not emotionally suited to women. They also have traditional attitudes about the role of women, especially those with children.

The women surveyed had doubts about their own ability and expected to be treated differently or unfairly. The views of both men and women seem stereotypical and need to be treated cautiously because the number participating in the focus group was very small.

One thing is certain. Irrespective of the discipline, women are not participating fully in science, engineering and technology. It is not clear why, and much study needs to be done on the subject. One contribution to the debate might be to excavate the historical record, not with the aim of politically correct revisionism, but with good scholarship that gives a fair evaluation of the intellectual

value of women's contribution, and puts their effort in the context of their time.

Giving societal context to their efforts would show the truly heroic nature of their actions. If women such as Caroline Herschel could be reclaimed from obscurity, attitudes might change. And if the stories of other women, such as Lisa Meitner and Irène Joliot-Curie were better known, no man would view women as unsuited to hard-edged competitive science.

Helen Gavaghan is a science and technology writer based in Hebden Bridge, West Yorkshire.

Useful contacts

4,000 years of women in science

♦ <http://www.astr.ua.edu/4000ws/4000ws.html>

On-line consultation by the Parliamentary Office of Science and Technology

♦ <http://www.mailbase.ac.uk/lists/hansard>

Nature's women in science information

♦ <http://helix.nature.com/debates>

The Committee of Vice-Chancellors and Principals' Athena project

♦ <http://www.athena.ic.ac.uk>

Women in Higher Education Register

♦ <http://www.where.ic.ac.uk>

Science, Engineering and Technology for Women Unit

♦ <http://www.set4women.gov.uk>

UK Association for Women in Science and Engineering

♦ <http://www.awise.org>

Royal Society of Chemistry

♦ <http://www.rsc.org/news.htm>

Daphne Jackson Memorial Fellowship for women returners

♦ <http://www.sst.ph.ic.ac.uk/trust/>

Women in Science and Engineering

♦ +44 (0) 20 7227 8421

Women returners network

♦ +44 (0) 20 7468 2290/1/2

New UK legislation aids fight against discrimination

With the deliberate, stately progress of an oil tanker changing course, the UK's universities are wresting the wheel round to steer a path away from practices that discriminate against women, minorities, part-time staff and those on fixed-term contracts. Progress might be slow, but it is under way.

In the case of contract research staff — postdocs on fixed-term contracts and part-time staff — new and pending legislation is providing a little extra impetus to the process of change. Legislation saying that part-time staff should get the same terms on a *pro rata* basis as full-time staff came into effect in April. This will have more impact on women than men because 44 per cent of women academics are part time, compared with only 10 per cent of male academics.

In the future, if someone is working five hours per week, the calculation of an equivalent *pro rata* salary would include preparation time, administration and marking, explains Tom Wilson, head of the Universities Department at the National Association of Further and Higher Education (NATFHE), one of the trade unions representing university lecturers. Until now, someone teaching five hours might have been paid for those five hours only. "Universities have been getting their teaching on the cheap," Wilson says.

Further legislation giving comparable rights to people on fixed-term contracts will be introduced in July 2001. The unions are currently negotiating an agreement that will say what this means in practice, and Wilson thinks this pending legislation is

concentrating the minds of universities on ways to improve the lot of staff on fixed-term contracts.

The legislation complements a move started in 1996 to improve conditions for contract research staff. Known as the Concordat, this is moving very slowly towards its goals: bringing terms and conditions in line with those of established staff; providing training and career guidance for contract research staff; and obtaining greater continuity of funding and employment when the research justifies it.

The Concordat was agreed by the UK research councils, the British Academy, the Royal Society, the Committee of Vice-Chancellors and Principals, the Standing Conference of Principals, and the Committee of Scottish Higher Education Principals. Progress towards its goals is monitored by the Research Careers Initiative, which published its most recent review in May. Sir Gareth Roberts, vice-chancellor of the University of Sheffield and chairman of the steering committee of the Research Careers Initiative, wrote in a preface to the review: "It is clear that the scale of change needs to be increased, and its pace accelerated."

The review found, for example, that compared with three years ago, more con-

tract research staff (53 per cent as opposed to 51 per cent) thought they were not on an equal footing with established staff when it came to university and departmental decisions. This perception persists despite seeming improvements in other aspects of the Concordat's aims, such as an increase in the number of staff that receive annual appraisals.

Bob Price, director of the Human and Corporate Resources Group of the Biotechnology and Biological Science Research Council, agrees that the universities need to ensure that the aims of the Concordat are met, but he does not think the new legislation covering fixed-term contracts will help. "All it will do is push up the bill for redundancy," he says. "Only a change in the way science is done so that an institution gets guaranteed funding for a longer time period, say ten years instead of the current three, would make a difference," he says.

H.G.

Further information

Research Careers Initiative

♦ <http://www.CVCP.ac.uk>

NATFHE

♦ <http://www.NATFHE.org>

The Concordat

♦ http://www.royalsoc.ac.uk/funding/fell_concord.htm

among minorities than the general population. Because of this, the National Institutes of Health (NIH) has been seeking ways to encourage minority participation in trials for more than 20 years.

Encouraging noises

But the problems clearly persist. Ruth Kirschstein, acting NIH director, has made fighting health disparity a primary goal of her year in the post. She hopes to find ways to encourage more research in diseases that disproportionately affect minorities, enrol more minorities in clinical trials, and encourage more minorities to go into research.

It is important to have minority practitioners to treat members of subpopulations and to enrol them in clinical trials, says Clifton Poodry, director of the Minority Opportunities in Research Division at the National Institute of General Medical Sciences. "Changing unhealthy patterns is harder if you don't look or talk like the people you're trying to help," he says. Outsiders can undermine their own best efforts through ignorance of local custom.

The need for balanced representation is even greater at the senior career level, where policy and funding decisions about minority health initiatives are made, but the scientist can arrive at this level only by progressing through graduate education and the lab. Sadly, the farther along the career trajectory, the thinner the representation of minority scientists.

How can more minority students be inspired to take up science studies that lead to science careers? Nurturing scientific potential in young people from non-mainstream backgrounds is a delicate challenge, says Poodry, who belongs to the Seneca Indian tribe of upstate New York.

Education in grammar and high school must be improved in the United States. "A student coming from a school that has a curriculum that's 40 years out of date may be lost when he gets to the modern research lab," says Poodry.

The old idea was that getting minority children into research labs would inspire them to become scientists, but exposure to research must come earlier, he continues.

US minorities stake their claim in science and engineering

In the North America, Blacks, Hispanics and Native Americans participate in the science and engineering labour force at a rate only one quarter that of the aggregate population. Hispanics account for more than a tenth of the total population, but hold only half a per cent of science and engineering PhDs (see diagram overleaf). Minority scientists and engineers have higher unemployment levels, are less likely to be tenured, have fewer publications, and get fewer federal grants and contracts than the majority white population.

The dismal portrait of minority participation in science painted by these figures — gathered from reports by the US National Science Foundation and Canadian National Research Council — signifies the waste of a huge pool of potential scientific talent. It also flouts the principle that science ought to be free of discrimination and open to everyone.

Obstacles to minority achievement in science block the way at every level from school to career. Deborah Jackson, an African American engineer at the Jet Propulsion Lab (JPL) in Pasadena, California, says the biggest problem minorities with science careers have is maintaining any gains they have made.

Since California's 1995 public law banning positive discrimination for minorities

in hiring and school admissions, the prospects for minority scientists in California have become increasingly bleak, Jackson says. "It's competitive at JPL for everyone, but if you're different, like me, it's even worse," she explains. "Early on I had lots of success, but right now I'm not feeling so great about where I've gone professionally."

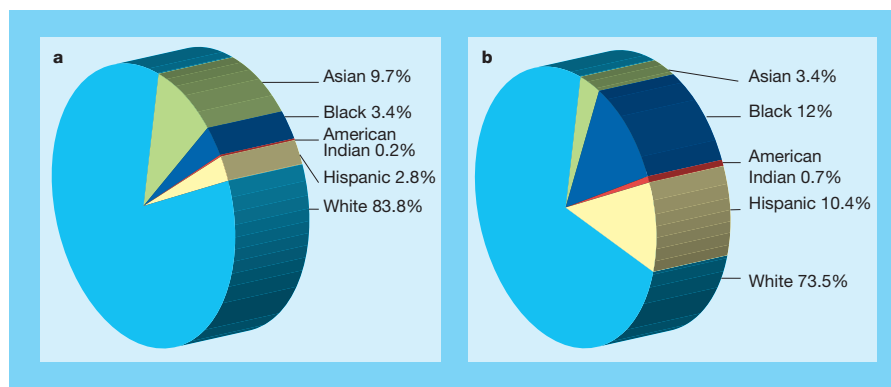
In the biomedical sciences, minority Americans are under-represented in clinical trials and are also scarce as investigators. Preventable and treatable diseases such as cardiovascular disease, asthma, diabetes, HIV/AIDS and lupus are more prevalent

Canadian aborigines get overlooked

Canada's roughly 1.2 million aboriginal people are ill-equipped for the future, says Merv Dewasha, president of the Canadian Aboriginal Science and Engineering Association, which is affiliated to Canada's National Research Council. In Canada, aboriginal people are 90 per cent under-represented in science and engineering professions.

Although the country is struggling with a skills shortage as it shifts from a commodity-based export economy to one based on technology, its aboriginal population is mostly overlooked. Seeing a long cycle of development ahead, Dewasha has focused on delivering this message to the Canadian government and to the tribal chiefs. The shortage of good maths and science teachers and the dispersion of tribal communities across the country's vast reaches present unusual difficulties, which distance learning, facilitated by CANARIE, Canada's high-speed Internet network, may mitigate.

P.W.



An uneven split. a, The proportion of scientists and engineers by race/ethnicity in 1995. b, The resident population of the United States by race/ethnicity in 1995.

Because research opportunities for undergraduates typically do not open up until the junior or senior year, populations such as the American Indian community, which have high early dropout rates, never get exposure to lab work in the first place.

Attempts to boost minority achievement in maths and science go back a generation. But the now largely dismantled programmes of the Affirmative Action era tended to have a built-in remedial bias that assumed the underperformance of non-whites.

Opinion today holds that a more productive approach is to push under-represented students rather than sequester them from the mainstream and feed them watered-down coursework. "High-school math programmes have been dumbed down for minorities, so students with talent are victimized by their teachers," says Manuel Berriozábal, who runs TexPREP, a maths and science enrichment programme for high-school students.

"If you push students to build communities based on a shared passion it's easier to help them make A's than to focus on their putative deficiencies," argues Uri Treisman, who started the Emerging Scholars Program at the University of Texas in 1988, based on his studies of minority achievement at Berkeley in the 1970s. The programme focuses on boosting students' grades in calculus.

Daunting numbers

Similar work needs to be done to help minorities settle into college. John Matsui runs the Biology Scholars Program at the University of California, Berkeley, which gives academic support to 400 mostly minority students. Part of the programme is Matsui's new course to help acclimatize transfer students from community colleges — an overlooked source of minority talent, he says — to university life. "I tell them about the culture of universities, how science is done and how it's funded, how courses are structured, why they may find assistant professors grumpy and hard to reach."

But for all the successes of such programmes, their graduates number only in the thousands, while the scope of the

problem must be counted in the millions. Antonio Flores is president of the Hispanic Association of Colleges and Universities, based in Texas, which represents 237 colleges and universities with high Hispanic student enrolments in North and Latin America. He cites studies that show Latino students lagging two years behind Whites in maths. Perhaps not surprisingly, a study last year by the Educational Testing Service shows that only 55 per cent of Latino eighth graders in the United States expect to go to college, compared with 64 per cent of African Americans and 72 per cent of Asians.

Much of the blame is attributable to below-par elementary and secondary schools in poorer districts. A current spate of lawsuits in California and other states, arguing that egregious deficiencies in academic programmes and physical facilities in poorer schools amount to a structural form of discrimination against minorities, aim to force state governments to guarantee basic resources in all schools. But with the inertia

of a generation of neglect, reform will come slowly.

Meanwhile, the shadows of an impending crisis grow longer. A recent report by President Bill Clinton's National Science and Technology Council predicts failing economic competitiveness and increasing social strains if present trends continue. And because the proportion of minority people in the United States will rise to 48 per cent over the next couple of generations, the disparity in representation will become even more pronounced unless something is done soon.

But the point might be made that minority under-representation and abysmal inner-city schools have been in place for at least a generation without harming the country's productivity and competitiveness. In the 1990s, the US economy roared past countries such as France and Japan, which have far better schools and early childhood education. So, leaving aside for a moment the question of social equality and limiting discussion to economic productivity, does science education really matter?

Of course it does. It is just that, as with many other essential resources, the United States imports it. Under the H1-B temporary visa plan for skilled workers, whose upper limit is likely soon to be increased to 200,000 a year, scientists and engineers flow into the United States from countries such as China, India, Taiwan and Canada.

"The quick fix is to go abroad and bring in those people to fill the jobs," says Antonio Flores. "But the strategic way of doing things would be to invest in the education of underprivileged young people here at home. It will improve not only our educational system, but our society as a whole."

Potter Wickware is a science writer in San Francisco.



Access to education provides a way out

The American Type Culture Collection (ATCC), a repository of bacterial and viral strains, cell culture lines and cDNA clones, has quite a diverse staff. About 32 per cent of its staff are from minority groups. But Yvonne Reid, who oversees cell culturing at the facility in Manassas, Virginia, worries that the number is not representative of minorities in science nationwide. Recently, Reid has noticed a decline in the number of African American scientists at meetings and academic labs at predominantly white institutions such as Harvard, Brown and Yale.

The sixth of seven siblings, Reid grew up in a rural district of Jamaica. Her father was a farmer and her mother stayed at home. "They could read and write, but were not educated people. It was my generation that broke that pattern," she says. During the 1950s and 1960s, when universal free education first became available in the country, the standardized curriculum meant that "the son or daughter of a labourer would have the same resources available in school as the son or daughter of a professor".

Reid earned a PhD in zoology from Howard University, Washington DC, in 1986. Her older sister has a doctorate in education from the Catholic University of America, Washington DC, another sister is a nurse, a cousin is a chemist with a degree from the University of the West Indies, and a nephew is doing a residency in surgery in New Jersey.

Reflecting on her own background, in which only education could lead the way out of poverty, Reid says: "The key to achievement, not only for minority persons but everyone, is high-quality, equal-opportunity education starting at an early age."

P. W.